°textA dissertation submitted in partial fulfillment of the requirements for the degree of

Certain observations concerning the effects of epistasis on complex traits and the evolution of genomes.

Maximilian Press

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2016

Reading Committee:

Christine Queitsch, Chair

Elhanan Borenstein, Chair

First committee member

Program Authorized to Offer Degree:
Department of Genome Sciences

University of Washington

Abstract

Certain observations concerning the effects of epistasis on complex traits and the evolution of genomes.

Maximilian Press

Co-Chairs of the Supervisory Committee:
Associate Professor Christine Queitsch
Department of Genome Sciences

Associate Professor Elhanan Borenstein
Department of Genome Sciences

The informational content of genomes is usually interpreted mimetically, which is to say, with a one-to-one relationship between genotypes at certain genomic positions and phenotypic outcomes. While such interpretations have the virtue of simplicity, they are empirically unsuccessful in elucidating the working of biological systems and in allowing us to predict variation in many phenotypes. Many have called for such models to explicitly consider epistasis, which can roughly be defined as any consideration of interactions between genomic elements (a semiotic rather than mimetic interpretation of genomic information). In this thesis, I consider some of the consequences of the existence of such epistasis. In the first part of this thesis, I consider a particular case of a fast-evolving genetic element (short tandem repeat, or microsatellite) that shows widespread epistasis, and propose that such elements are likely to accumulate epistatic interactions by acting as mutational modifiers. I go on to show molecular mechanisms by which the element participates in epistasis, their phenotypic consequences, and make some observations on other short tandem repeats. In the second part of this thesis, I start with the assumption of epistasis between genes, and explore how this assumption can be used to understand the evolution of bacterial genome con-

tent. First, I take Hsp90, the known epistatic hub, and infer its coevolution with other genes through coordinated gains and losses across bacterial diversity. I further extend the underlying phylogenetic model to predict new "clients" of bacterial Hsp90, which have remained elusive when pursued through purely experimental approaches. Collaborators were able to validate certain of these predicted clients. Last, I attempt an analogy between prokaryotic genome evolution and the much better-understood field of protein evolution. I propose that, like protein evolution by substitution, genome evolution by horizontal acquisition of genes is substantially constrained by epistasis. I go on to infer the existence of such epistatic dependencies, where one gene in an ancestral genome promotes the acquisition of a second gene. A network of such dependencies shows a chronological structuring of gene acquisitions through prokaryotic evolution, suggesting universal assembly patterns by which genomes acquire functions. I go on to show that these dependencies are taxonomically universal (i.e. not restricted to particular phyla), and that they are sufficient to make reasonably good predictions about what genes a genome will gain in the future. This predictability of genome evolution by horizontal transfer confirms a major assertion of the protein evolutionists, that constraining epistasis leads to predictable evolutionary outcomes. Together, these observations indicate that the genetic architecture of traits and the content of genomes are shaped by the existence of networks of epistasis, reflecting the complex wiring of underlying biological functions.

TABLE OF CONTENTS

LIST OF FIGURES

Figure Number Page

ACKNOWLEDGMENTS

I would like to thank my advisors, Christine Queitsch and Elhanan Borenstein, for letting me ride my ideas as far as I did (and equally, for curtailing those ideas when they got ridiculous). I would also like to thank Bob Kaplan, Katie Peichel, and Sue Biggins for saving Elhanan and Christine a world of grief in mentoring me before I started my doctoral work.

I would like to thank my thesis committee: Joe Felsenstein, Willie Swanson, and Evgeni Sokurenko. Joe in particular was generous with his time early on, in helping me to understand phylogenies and discrete character evolution.

I would like to thank my co-belligerents in the Queitsch and the Borenstein labs, for what was surely a miracle of forbearance.

I would like to thank my parents for everything.

I would like to thank everyone else, in whose enumeration we could easily exhaust ourselves.

And Sarah.

"I will ask you to mark again that rather typical feature of the development of our subject; how so much progress depends on the interplay of techniques, discoveries and new ideas, probably in that order of decreasing importance."

Sydney Brenner

That generation's dream, aviled
In the mud, in Monday's dirty light,

That's it, the only dream they knew,
Time in its final block, not time

To come, a wrangling of two dreams.
Here is the bread of time to come,

Here is its actual stone. The bread
Will be our bread, the stone will be

Our bed and we shall sleep by night.
We shall forget by day, except

The moments when we choose to play
the imagined pine, the imagined jay.

Wallace Stevens

DEDICATION

to my Sarah

Chapter 1

INTRODUCTION

1.1 Practical and philosophical approaches to the problem of genetic architecture

[4]

1.2 Linear models and Fisherian quantitative genetics

1.3 Epistasis

1.4 Fast-evolving genetic elements

1.4.1 Short tandem repeats (STRs)

1.4.2 Horizontal acquisition of genetic material

Chapter 2

BACKGROUND-DEPENDENT EFFECTS OF POLYGLUTAMINE VARIATION IN THE *ARABIDOPSIS THALIANA* GENE *ELF3*

A version of this chapter was published under the following reference:

Soledad F. Undurraga, Maximilian O. Press, Matthieu Legendre, Nora Bujdoso, Jacob Bale, Hui Wang, Seth J. Davis, Kevin J. Verstrepen, and Christine Queitsch. Background-dependent effects of polyglutamine variation in the *Arabidopsis thaliana* gene *ELF3*. Proceedings of the National Academy of Sciences of the United States of America, 109(47):19363-7, November 2012.

Soledad Undurraga, Jacob Bale, Nora Bujdoso, and Hui Wang generated transgenic lines, performed experiments, and contributed figures.

Supplementary figures and tables can be found in Appendix A.

2.1 Summary

Tandem repeats (TRs) have extremely high mutation rates and are often considered to be neutrally evolving DNA. However, in coding regions, TR copy number mutations can significantly affect phenotype and may facilitate rapid adaptation to new environments. In several human genes, TR copy number mutations that expand polyglutamine (polyQ) tracts beyond a certain threshold cause incurable neurodegenerative diseases. PolyQ-containing proteins exist at a considerable frequency in eukaryotes, yet the phenotypic consequences of natural variation in polyQ tracts that are not associated with disease remain largely unknown. Here, we use Arabidopsis thaliana to dissect the phenotypic consequences of natural variation in the polyQ tract encoded

by *EARLY FLOWERING 3* (*ELF3*), a key developmental gene. Changing ELF3 polyQ tract length affected complex ELF3-dependent phenotypes in a striking and non-linear manner. Some natural ELF3 polyQ variants phenocopied elf3-loss-function mutants in a common reference background, although they are functional in their native genetic backgrounds. To test the existence of background-specific modifiers, we compared the phenotypic effects of ELF3 polyQ variants between two divergent backgrounds, Col and Ws, and found dramatic differences. Our data support a model in which variable polyQ tracts drive adaptation to internal genetic environments.

2.2 Introduction

In coding regions, tandem repeat (TR) copy number variation can have profound phenotypic effects [2]. For example, TR copy number mutations that expand polyglutamine (polyQ) tracts past a threshold number of glutamines can cause incurable neurodegenerative diseases such as Huntington's disease and Spinocerebellar Ataxias [1, 3]. PolyQ tract length correlates with onset and severity of polyQ expansion disorders, but for intermediate polyQ tracts this correlation is far weaker (4-8), suggesting the existence of genetic and environmental modifiers (9-12). Despite their potential for pathogenicity, variable polyQ tracts occur frequently in eukaryotic proteins, many of them functioning in development and transcription (1, 13-15). Model organism studies have suggested that coding TRs are an important source of quantitative genetic variation that facilitates evolutionary adaptation (1, 16-19). For example, TR copy number variation in the yeast gene FLO1 correlates linearly with flocculation (20), a phenotype that is important for stress survival (17). As polyQ tracts often mediate protein interactions (2, 3, 21), polyQ-encoding TR copy number mutations could produce large and possibly adaptive phenotypic shifts. To determine the phenotypic impact of naturally occurring polyQ variation (18, 22, 23) in a genetically tractable model, we focused on the gene ELF3, which encodes a polyQ tract that is highly variable across divergent Arabidopsis thaliana strains (accessions) (19, 24). ELF3 is a core component of the cir-

cadian clock and a potent repressor of flowering, and is considered a "hub protein" for its many interactions with various proteins (24-31). Consequently, elf3 loss-of-function mutants show pleiotropic phenotypes: they flower early, show poor circadian function, and grow long embryonic stems (hypocotyls) in light (25-27, 29, 30, 32). Single nucleotide polymorphisms (SNPs) in ELF3 affect shade avoidance, a fitness-relevant plant trait (24, 33). ELF3 polyQ variation has been suggested to correlate with two parameters of the circadian clock, period and phase (19). The ELF3 polyQ tract may mediate ELF3 membership in protein complexes, though thus far no ELF3-binding protein is known to bind it (26, 28-30). We discovered that altering polyQ tract length has dramatic effects on ELF3-dependent phenotypes and that these effects are dependent on genetic background.

2.3 Methods

2.3.1 Plant Materials and Growth Conditions.

The 181 Arabidopsis thaliana accessions are as previously described (1). The loss-of-function EARLY FLOWERING 3 (elf3) mutants are: (i) *elf3-4*, containing a CCR2::LUC transgene (ecotype Ws) (2, 3) and 2) *elf3-200*, the GABI750E02 T-DNA insertion mutant (ecotype Col-0) (4). For hypocotyl experiments, seeds were sterilized with Ethanol and plated onto 1x Murashige and Skoog (MS) basal salt medium supplemented with 1x MS vitamins, 1% sucrose, 0.05% Mes (wt/vol), and 0.24% (wt/vol) phytagel. After stratification in the dark at 4° C for 3 d, plates were transferred to an incubator (Conviron) that was set to either short day (SD) (8L:16D at 20° C) or long day (LD) (16L:8D at 22° C : 20° C), with light supplied at $100 \mu mol * m^2 * s^{-1}$ by cool-white fluorescent bulbs. For growth on soil, seeds were stratified at 4° C for 3 d, and then grown in Sunshine #4 soil under cool-white fluorescent light at either LD or SD at 20 °C. Seedlings used for RNA extractions were grown on soil under LD conditions and harvested on day 10. Samples for ELF3 expression measurements were collected at Zeitgeber time (ZT) 20. Samples

for Phytochrome- interacting Factor 5 (PIF5) expression measurements were collected a ZT 8. Samples for and Pseudoresponse regulator 9 (PRR9) expression measurements were collected at ZT 0, 5, and 8.

2.3.2 Generation of ELF3 Transgenic Plants.

To generate A. thaliana transgenics carrying different ELF3 tandem repeat (TR) alleles, the cDNA clone RAFL09-28-E05 (RIKEN BRC) (5, 6), containing the ELF3 coding region and 3' UTR (Col-0 accession) was used. This cDNA clone lacks the small 5' intron. Two re- striction sites, Nar1 and Nco1, were inserted into the ELF3 coding sequence using the QuikChange Site-Directed Muta- genesis kit (Stratagene) (primer information in Table A5). The polyglutamine (polyQ)-encoding region was amplified from ac- cessions containing selected TR copy number alleles (primer information in Table A5, TR allele information in Table A1). These PCR products were digested with Nar1/Nco1 and ligated into the previously mutagenized ELF3 coding region. An artifi- cial allele lacking the TR was generated by site-directed muta- genesis (primer informa- tion in Table A5). Mutated plasmids and all ligation products were sequenced to ensure accuracy. The ELF3 alleles were cloned into pENTR1A (Invitrogen). A 2-kbp NotI fragment containing the ELF3 promoter was inserted up- stream of each ELF3 cod- ing sequence. The fragments containing the ELF3 promoter, ELF3 coding sequence, and the ELF3 3' UTR were recombined using Gateway LR Clonase II (In- vitrogen) into a modified pB7WG2 (7), which lacks the CaMV- 35S promoter. The region en- coding the polyQ tract of each construct was sequenced to ensure accurate TR copy number. The plasmids were used to transform Agrobacterium tumefaciens GV3101. Subsequently, Arabidopsis elf3 mutants were trans- formed by the flower dip method (8). Transformants were se- lected on Basta (Liberty herbicide; Bayer Crop Science) and propagated for three to four generations. The accuracy of the transgenes was confirmed by PCR (primer information in Table A5). All Ws phenotypic assays were performed in homozygous transgenic plants with expression levels between 0.8- and 4.5- times the

respective ELF3 wild-type (Figure A1C); for Col lines, transgene expression levels were between 0.3- and 4.3-times the respective ELF3 wild-type (Figure A1D). Analyzed plant lines are in Tables A2-A4.

2.3.3 RNA Extractions and Real-Time PCR.

Total RNA was extracted from 30-mg frozen tissue using the SV Total RNA Isolation System (Promega). Subsequently, 2 μg of RNA were subjected to DNase treatment using Ambion Turbo DNA-free Kit (Applied Biosystems). RNA integrity and purity were checked with an Agilent Bioanalyzer using the RNA 6000 Nano Kit (Agilent Technologies). For cDNA synthesis, 200 ng of DNase-treated RNA was reverse-transcribed using the Transcriptor First Strand cDNA Synthesis Kit (Roche) and oligo dT primers. Transcript abundance was determined by real-time quantitative PCR using the LightCycler 480 system (Roche), with LightCycler 480 SYBR Green I Master (Roche) and the following PCR conditions: 5 min at 95 °C, followed by 35 cycles of 15 s at 95 °C, 20 s at 55 °C, and 20 s at 72 °C. To ensure that PCR products were unique, a melting- curve analysis was performed after the amplification. UBC21 expression (At5g25760) was used as a reference. All quantitative RT- PCR primers were designed with the LightCycler Probe Design Software (Roche). Sequences for real-time PCR primers are shown in Table A6. Relative quantification was determined with the $\Delta\Delta C_T$ Method (9). Error was calculated as previously described (10).

2.3.4 Thermal Asymmetric Interlaced PCR.

High-efficiency thermal asymmetric interlaced (TAIL)-PCR was performed as previously described (11) to obtain the flanking sequence of the construct in- tegration site (left border). Briefly, a preamplification step was performed with primers LAD and LB-0a (Table A7), followed by primary TAIL-PCR with primers AC1 (11) and LB-1a (Table A7), and 1 μL of a 1/40 dilution of the preamplification product as a template. A secondary TAIL-PCR with primers AC2 (11) and LB-2a (Table A7) was performed

with 1 μL of a 1/10 dilution of the primary TAIL-PCR product. Next, 3-kbp products were extracted from agarose gels and subsequently Sanger-sequenced. Only sequences containing the T-DNA left border were considered.

2.3.5 Developmental Phenotype Assays.

For measurements of hypocotyl length, seedlings were grown on vertical plates for 15 d in a pseudorandomized design under either SD or LD conditions (12). Hypocotyl length was measured with ImageJ on digital images (http://rsbweb.nih.gov/ij/). For measurement of flowering time, seeds were planted in sheet pots (36 pots per tray) in a randomized design and trays were rotated daily. Flowering time was re- corded as the day when the inflorescence reached 1 cm in height. Rosette leaf number was determined on the same day. Petiole- length/leaf-length (PL/LL) ratio for leaf four was determined on day 45. Least-square means for all traits were derived from a linear regression analysis for each trait separately. ELF3 TR copy number was modeled as a nominal variable and independent transgenic lines carrying the same ELF3 TR allele were analyzed together. We tested for significant phenotypic differences con- ferred by the different ELF3 TR alleles by using Tukey-HSD tests with $\alpha = 0.05$ that accommodate nonnormal data.

2.3.6 Luciferase Imaging and Period Analysis.

Luciferase assays were performed with lines containing the CCR2::LUC reporter. Seeds were surface sterilized with a 70% (vol/vol) ethanol wash followed by a second wash with 33% (vol/vol) Klorix with Triton X-100, and then rinsed twice with sterile water. Seeds were plated on MS3 medium [pH 5.7, 3% (wt/vol) sucrose, 1.5% (wt/vol) PhytoAgar, and 15 μg/mL hygromycin B]. They were subsequently stratified for 4 d at 4 °C in the dark and entrained under 12-h light:12-h dark cycles under white fluorescent light ($\sim 10 \mu mol * m^{-2} * s^{1}$) at 22 °C. On the sixth day, a minimum of 24 seedlings per line was transferred to 96-well TopCount (Perkin-Elmer) plates containing 200 mg MS3 agar. We added 5 mM Luciferin in 0.01% Triton X-100 and entrained seedlings for

another cycle before luminescence was detected using a Packard/Perkin-Elmer TopCount Scintillation and Luminescence Counter. Red and blue light-emitting diodes ($100\mu mol * m^{-2} * s^{-1}$) were used as a light source during this analysis. During the first 24 h of luminescence detection, plants were grown in 12-h light:12-h dark and then released under constant light conditions to measure the free- running period. Each individual was measured approximately every 30 min for a minimum of 5 d. Luminescence levels were quantified and analyzed as previously described (2, 3) using the macro suites TopTempII and Biological Rhythms Analysis Software System (13). Period length and relative amplitude error (RAE) were estimated using fast Fourier transform nonlinear least squares (14). Period values scored with RAE values below 0.4 were considered robustly rhythmic (15).

2.3.7 Principal Component Analysis.

We clustered our phenotypic data using principal component analysis (PCA) to find patterns corresponding to genotypes. We excluded the phenotype of rosette leaf number in SD, for which data were missing for several alleles. The phenotypes included in the analysis are: Days to flowering in SD and LD conditions, hypocotyl length under SD and LD PL/LL for the fourth leaf in SD, and rosette leaf number in LD. For analyses involving Col lines, the SD PL/LL ratio phenotype was omitted because of lack of data, and PCA was thus based on the remaining five phenotypic variables. For each phenotype in each genetic background (either Ws or Col-0), we calculated the mean phenotype of the independently generated lines for each *ELF3-TR* allele, giving us a 28 x 6 matrix of mean phenotypes for the 28 genotypes for each of six phenotypic variables. Within each background, we ranked the genotypes for each phenotype. Ranks were transformed into a standard normal distribution based on their percentile, using the R function qnorm. Using this transformed dataset, we performed PCA using the R function *prcomp* (R Foundation for Statistical Computing, http://www.r-project.org/, 2011). We performed PCA for each background separately, and then for both backgrounds

together. Rank-normalization was necessary to compare (i) phenotypes measured on different scales and (ii) Ws- and Col-derived plants, between which backgrounds absolute phenotypic differences exist. Consequently, the rank-normalization increases stability of our estimates, as our dataset is relatively small and PCA□ as assumptions of normality were not met by our raw dataset. PCA on raw values scaled to a standard normal distribution gave similar results. Biplots were generated with the R $biplot$ function on $prcomp$ function output.

2.4 ELF3-TR-variation affects ELF3-dependent phenotypes.

Among 181 natural A. thaliana accessions, the *ELF3-TR* encoded between 7 and 29Q (Table A1, Figure A1a). For comparison, polyQ expansions over 20Q are associated with disease in the context of the SCA6 gene, though most other disease-associated polyQ expansions are longer (2, 19, 24). The most frequent *ELF3-TR* encoded 16Q, whereas the shortest TR (7Q) was found in the reference strain Col-0. We set out to test whether naturally occurring *ELF3-TR* alleles affect ELF3-dependent phenotypes and whether they do so in a linear manner as suggested by association studies (19) and found for coding TR variation in other genes (16, 20). We generated expression-matched transgenic lines for most natural *ELF3-TR* alleles in the loss-of-function *elf3-4* mutant (Ws background, Table A2a, Figure A1c) (32) and measured their flowering time and circadian clock-related phenotypes (Figures 1, SAa-g). *ELF3-TR* variation significantly affected ELF3-dependent phenotypes, but there was no evidence of a linear relationship. The different *ELF3-TR* alleles resulted in phenotypes ranging from nearly full complementation of *elf3-4* to nearly phenocopying the loss-of-function mutant. We used principal components analysis to describe the complex effects of *ELF3-TR* alleles on all tested ELF3-dependent phenotypes (PCA, Figures 1a, A2h-j). Principal component 1 (PC1) corresponds to general functionality of ELF3 in all measured phenotypes, with wild-type Ws and mutant *elf3-4* defining the extremes. Separation along PC1 is driven by the tendency of plants with functional ELF3 to show short hypocotyls, late

flowering, increased rosette leaf number, and short petioles (Figures 1b-d, A2). The endogenous ELF3-16Q allele complemented both the early-flowering and long-hypocotyl phenotypes of *elf3-4* (Figures 1b-d, A2). In contrast, both the long ELF3-23Q and the short ELF3-7Q allele (endogenous TR alleles in Br-0/Bur-0 and Col-0, respectively) behaved similarly to the *elf3-4* loss-of-function allele (Figures 1b-d, A2), although they are functional in their native backgrounds. Neither Col-0 nor Br-0 and Bur-0 show the phenotypic characteristics of elf3-mutants (early flowering (34), long hypocotyls (35) and long petioles (36)), suggesting that *ELF3-TR* alleles may interact with background-specific modifiers. ELF3-0Q, an artificial ELF3 allele lacking the TR, partially complemented *elf3-4* (Figures 1a, S2). Hence, the polyQ-encoding TR is not necessary for all ELF3 function, but changes in TR copy number are sufficient to enhance or ablate ELF3 function.

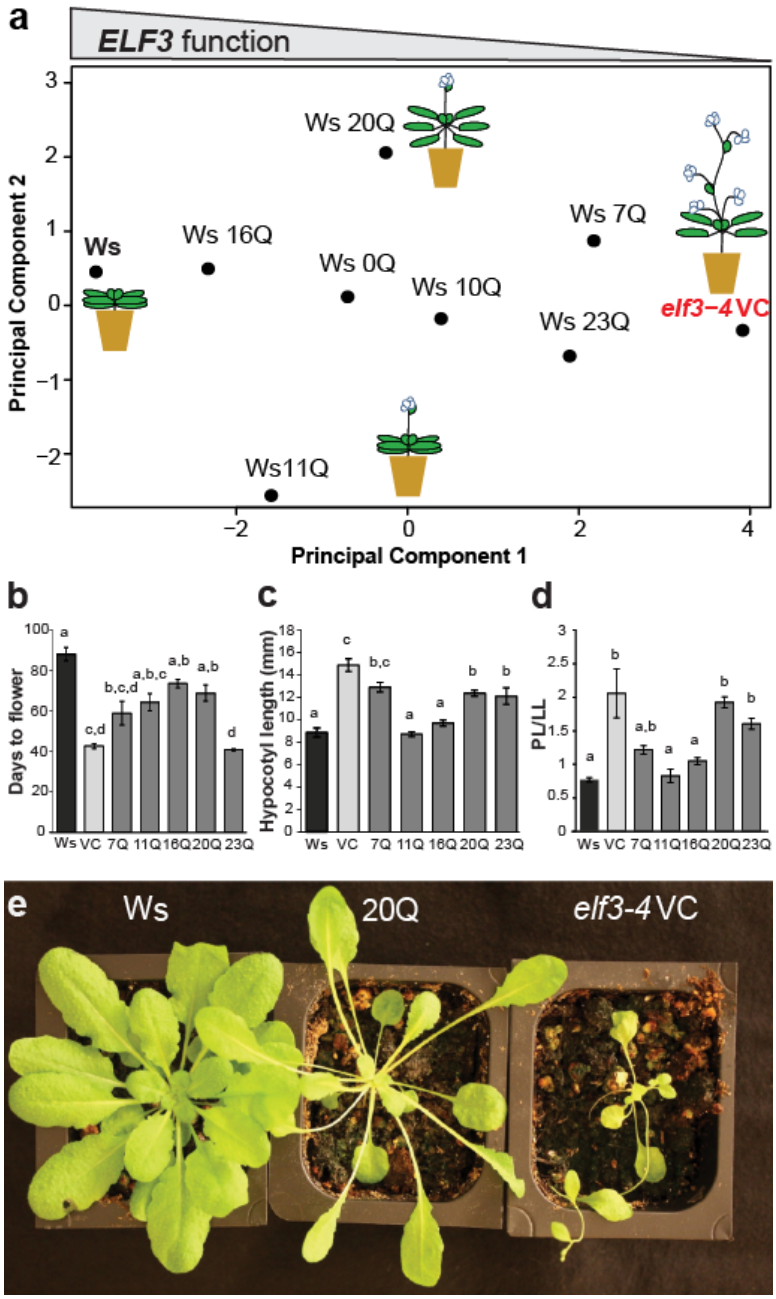

a
ELF3 function
Principal Component 2
Principal Component 1
Ws
Ws 16Q
Ws 0Q
Ws 10Q
Ws11Q
Ws 20Q
Ws 7Q
Ws 23Q
elf3−4 VC
b
Days to flower
c
Hypocotyl length (mm)
d
PL/LL
Ws VC 7Q 11Q 16Q 20Q 23Q
e
Ws
20Q
elf3-4 VC

Figure 2.1

Figure 2.1: ELF3-TR variation has nonlinear phenotypic effects. (A) PCA of developmental traits of all ELF3-TR copy number variants. A. thaliana images illustrate ELF3-TR effects on the traits days to flower and hypocotyl length under SD and LD, petiole-length/leaf-length ratio (PL/LL) under SD only, and rosette leaf number under LD only. The contributions of specific phenotypes to PCs are in Figure A2J. Representative TR copy number alleles are shown from an analysis including all alleles (for all alleles see Figure A2 H and I). (B) Days to flower under SD conditions for selected lines. n = 6 plants per transgenic line. (C) Hypocotyl length at 15 d under SD for selected lines. n = 20-30 seedlings per transgenic line. (D) PL/LL of the fourth leaf for selected lines. Data are from the same plants as in B. (E) Plants carrying the ELF3-20Q allele (Center) are specific hypomorphs under SD with the elongated petioles of the elf3-4 mutant (vector control, VC, Right) and a wild-type flowering phenotype (Ws, Left). ELF3-TR alleles are indicated with the number of Qs encoded, Ws is wild-type, VC is the elf3-4 vector control. Error bars are SEMs. Genotypes labeled with different letters differed significantly in phenotype by Tukey's HSD test. For all Ws-background phenotype data, see Figure A2 A□G. Data are from multiple independently generated expression-matched (Figure A1C) T3 and T4 lines for each TR copy number allele (Table A2). These experiments were repeated at least once with similar results. The tested ELF3-20Q lines contained unique insertions that did not affect genes with known function (Table A3).

PC2 separated ELF3-20Q and ELF3-11Q, which behaved as hypomorphs in certain phenotypes but not others (Figure 1a). For example, ELF3-20Q plants had significantly longer hypocotyls than wild-type and its petioles phenocopied the extremely long petioles of the *elf3-4* mutant (Figure 1c-e), but they did not differ from wild-type in flowering time (days to flower, Figure 1b). The existence of both general and specific hypomorphs suggests that polyQ variation affects the multiple ELF3 functions separately. As part of a protein complex, ELF3 affects expression of Phytochrome-interacting Factor 5 (PIF5) and Pseudo-response regulator 9 (PRR9) (28, 37, 38). PIF5 and PRR9 expression were

strongly affected by ELF3 polyQ variation (Figure A3). ELF3-16Q phenocopied wild-type PRR9 and PIF5 expression, and the hypomorphic ELF3-23Q phenocopied *elf3-4* (28, 37, 38), mirroring their developmental phenotypes. Consistent with their divergence along PC2 (Figure 1a), ELF3-11Q and ELF3-20Q differed in their effect on PRR9 expression, but not on PIF5 expression (Figure A3a, b), demonstrating that ELF3 polyQ variation differentially affects the regulation of downstream genes.

2.5 ELF3-TR*-variation modulates the precision of the circadian clock*

To directly assess the role of ELF3 polyQ variation in the circadian clock, we used the CCR2::LUC reporter system (25, 39). We observed little difference in circadian period among wild-type Ws and tested ELF3-TR alleles (Figure A4a), contradicting a previously observed association of TR copy number with period in natural accessions (19). However, we found that the relative amplitude error (RAE) of oscillation varies substantially across ELF3-TR genotypes (Figures 2a, S4b). RAE measures the precision of a circadian period (40): high RAE values (> 0.4) indicate poor oscillation and clock dysfunction (41). The endogenous Ws ELF3-16Q nearly complemented the elf3-4 RAE defect, whereas the TR alleles ELF3-7Q, ELF3-10Q, and ELF3-23Q showed higher RAE, approaching arrhythmic *elf3-4* levels (Figure 2a, b), consistent with their hypomorphic performance in other ELF3 traits (close to *elf3-4* in PC1, Figure 1a). Together, these results suggest that ELF3 polyQ tract length is a critical determinant of circadian clock precision, but not period length, in A. thaliana.

2.6 ELF3-TR-variation interacts with genetic background.

To test our hypothesis that *ELF3-TR* variation interacts with genetic background, we regenerated all *ELF3-TR* transgenic lines in the *elf3-200* loss-of-function mutant with matched transgene expression (Col background, Table A2c, Figure A1d) (42). We used PCA to compare *ELF3-TR* effects between Ws and Col backgrounds (Figures 3a, S5). The Col-specific ELF3-7Q allele complemented *elf3-200* in some traits such as flower-

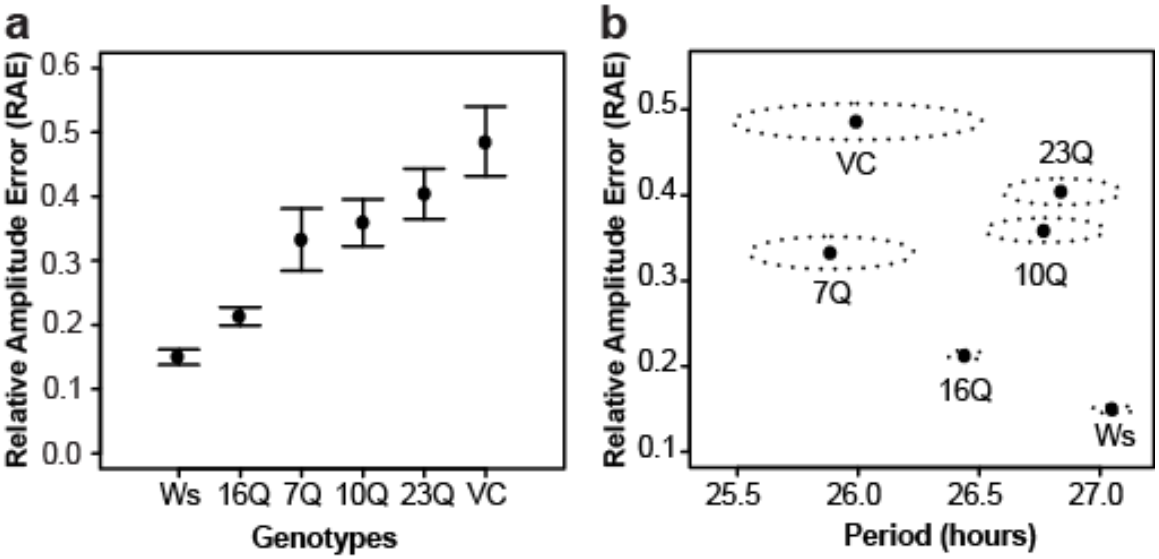


Figure 2.2: ELF3-TR variation modulates the precision of the circadian clock. (A) RAE of CCR2::LUC circadian oscillation in seedlings with indicated ELF3-TR alleles. Bars represent 99% confidence intervals. (B) Mean values of circadian period and RAE (points) were measured in seedlings with indicated ELF3-TR alleles. Dotted ellipses represent SEMs for both period and RAE. Note that plants with high RAE have extremely unreliable estimates of circadian period. Biolumines- cence rhythms from the CCR2::LUC reporter in ELF3-TR transgenic lines were used to measure circadian parameters under LL after 5 d of entrainment in 12-h light:12-h dark cycles. n≥100 seedlings for all geno- types. Aggregate data from four independent experi- ments are shown. See Fig. A4 for RAE and period data for all alleles.

ing time (in short days, SD) and hypocotyl length (in long days, LD), but not others (Figures 3a, b, S5, S6). This result may be due to the absence of the small 5' intron from the ELF3 construct used in this study. However, there was still a dramatic spread of phenotypes: all longer *ELF3-TR* alleles (>20 Qs) nearly complemented *elf3-200*, delaying flowering and shortening hypocotyls, whereas few of the shorter alleles did (Figures 3, S5, S6). Results were similar when the Col data were analyzed alone (Figure A6). Thus, in contrast to our results in the Ws background, *ELF3-TR*s appeared to show a threshold effect for TR copy number in the Col background. We speculate that the intensive laboratory propagation of the Col-0 accession may have altered selection on the *ELF3-TR*, resulting in an extremely short "hypomorphic" allele, whereas under natural conditions a longer TR might be more functional. Comparing TR allele effects between the two backgrounds revealed striking differences. For example, the ELF3-23Q allele was a general hypomorph in the Ws background (*elf3-4*), whereas it produced highly functional ELF3 in the Col background (*elf3-200*, Figure 3). In turn, the ELF3-16Q allele produced highly functional ELF3 in the Ws background (*elf3-4*), but was a general hypomorph in the Col background (*elf3-200*). The consistent performance of the artificial ELF3-0Q allele across backgrounds suggests that the background effect is TR-dependent (Figures 3a, S5). Collectively, our results support that *ELF3-TR* alleles interact with background-specific modifiers.

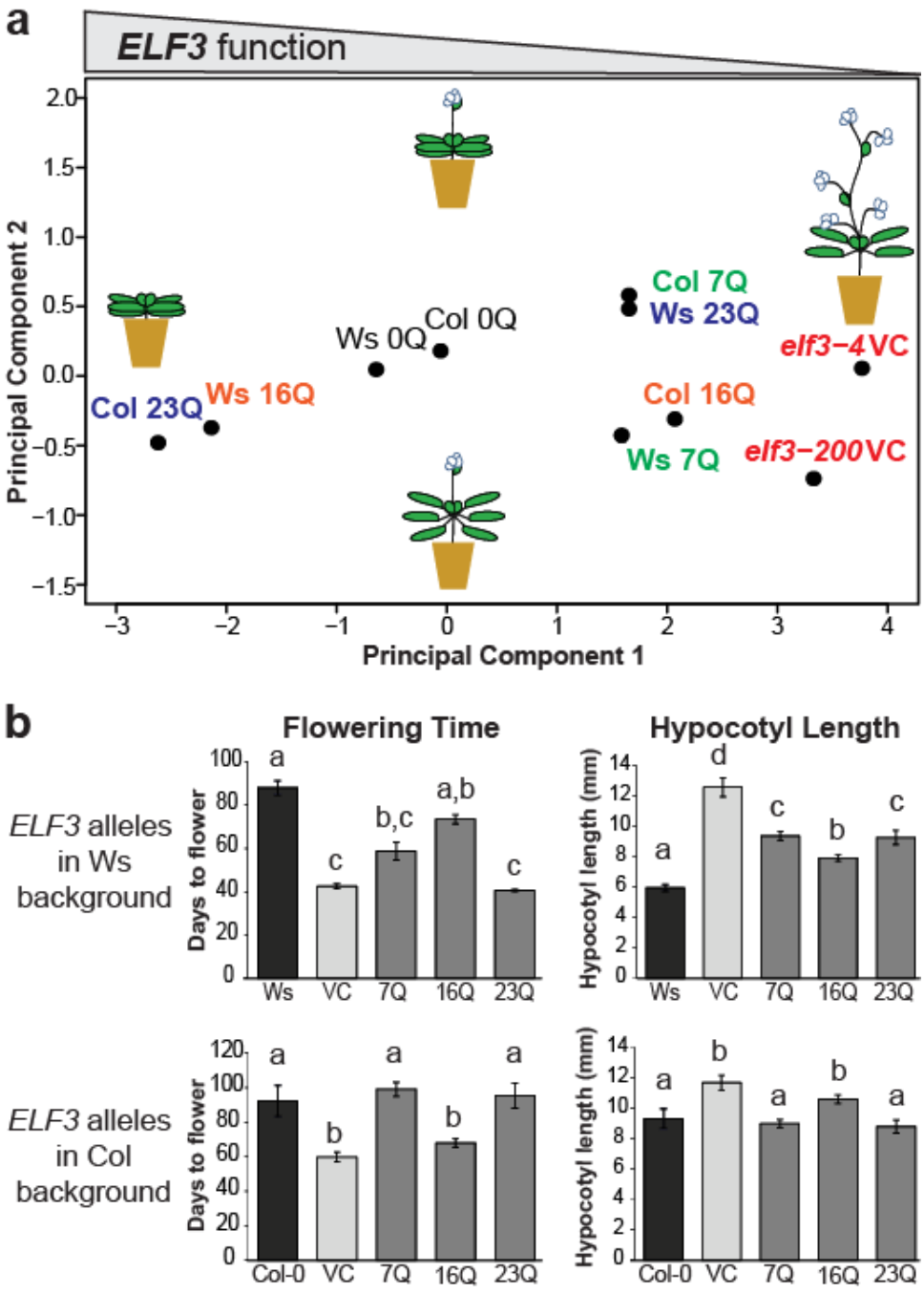

a
ELF3 function
Principal Component 2
Principal Component 1
Col 7Q
Ws 23Q
Col 0Q
Ws 0Q
elf3−4VC
Col 23Q
Ws 16Q
Col 16Q
Ws 7Q
elf3−200VC
b
Flowering Time
Hypocotyl Length
ELF3 alleles in Ws background
ELF3 alleles in Col background
Days to flower
Hypocotyl length (mm)
Ws
VC
7Q
16Q
23Q
Col-0

Figure 2.3

Figure 2.3: The phenotypic effects of ELF3-TR variation are strongly background-dependent. (A) PCA of developmental traits of all ELF3-TR alleles in Ws and Col genetic backgrounds. Shared color indicates a given ELF3-TR allele in both genetic backgrounds. A. thaliana images are as in Fig. 1A. The contributions of phenotypes to principal components are similar to Fig. 1A, except that PC2 is inverted (no effect on interpretation, loadings in Fig. A5C). Representative TR copy number alleles are shown from an analysis including all alleles (for all alleles see Fig. A5; for Col-background specific PCA, see Fig. A6). (B) Days to flower under SD and hypocotyl length under LD differ for particular TR alleles between Ws (Upper) and Col (Lower) backgrounds. ELF3-TR alleles are indicated with the number of Qs encoded, Ws and Col-0 are wild-type, elf3-4 and elf3-200 are respective vector controls (VC). Error bars represent SEM. Genotypes labeled with different letters differed significantly in phenotype by Tukey-HSD test. For all Col-background phenotype data, see Fig. S6 A□G. Data are from multiple independently generated expression-matched (Fig. A1C and D) T3 and T4 lines for each TR copy number allele (Table A4). These experiments were repeated at least once with similar results.

2.7 Col ELF3 *allele is not haploinsufficient in Col x Ws hybrids.*

To address whether Ws and Col-specific background effects are sufficient for altered hybrid phenotypes, we generated F1 populations between wild-type and elf3 null plants in the Ws and Col backgrounds and measured ELF3 function by assessing hypocotyl length. Ws x Col F1 hybrids resembled their wild-type parents (Figure 4). F1 hybrids containing both loss-of-function alleles had significantly longer hypocotyls than either parent (Figure 4). Both ELF3 alleles were haplosufficient in F1 crosses within their native backgrounds, as expected for recessive mutants (Figure 4). In stark contrast, we observe that ELF3-Col, but not ELF3-Ws, phenocopied the extreme hypocotyl length of the double loss-of-function mutant (Figure 4). Consistent with the results from our

transgenic lines, our F1 hybrid data suggest that full ELF3 function depends on a permissive genetic background.

Unfortunately, propagation of these F_1 hybrids to the F_2 generation and subsequent Col x Ws crosses revealed that the data in Figure 4 do not generalize to other crosses, and probably represent a spontaneous mutation in the Col background leading to ELF3 inactivation. In the face of such equivocal evidence, we suggest that more intensive genetic or biochemical experiments will be necessary to determine the relevant background modifiers of *ELF3-TR* variation. For two such approaches, refer to Chapter 5.

2.8 Discussion

Our results demonstrate that natural ELF3 polyQ variation that is not associated with disease has dramatic phenotypic consequences, and that these consequences depend on genetic background. For ELF3, in at least the Ws background, the relationship between TR copy number and phenotype does not follow a linear or threshold pattern as observed for other coding TR and polyQ disorders (1, 2, 16, 17, 20). Studies correlating TR variation with phenotype often apply linear models, treating TR copy number as a quantitative variable (19, 22, 23). Our data show that this approach is not appropriate for all TRs. Instead, *ELF3-TR* alleles seem "matched" to specific genetic backgrounds, in which they are functional, whereas they are incompatible with other backgrounds. The haploinsufficiency of the elf3-col allele in Ws x Col hybrids supports this interpretation. In contrast, the ELF3-Ws allele is haplosufficient in hybrids, indicating that the ELF3-Col x Ws incompatibility is asymmetric. This observation agrees with Orr's assertion that incompatibility between recently diverged populations is usually asymmetrical, because it tends to arise from the derived allele (i.e. ELF3-Col) (43). Variable TRs, and the *ELF3-TR* in particular, have been previously suggested as agents of adaptation to new external environments (1, 16, 17, 20, 24, 44). Our results suggest that polyQ-encoding TRs are also agents of coadaptation within genomes. We speculate

that the observed background effects arise from background-specific polymorphisms in genes encoding physically interacting proteins (26, 28-30). TRs have a far higher mutation rate than non-repeated regions (10-4 per site per generation for TR vs. 10-8 for SNPs) (45, 46) and, as we show, their expansion or contraction can have dramatic phenotypic impact. ELF3's partner proteins may have acquired compensatory mutations to accommodate new *ELF3-TR* variants and vice versa. Alternative explanations for the background effects are compensatory mutations in ELF3 (intragenic suppressors), or ELF3 interactions that are unique to a given background. Intragenic variation and protein modification can play an important role in polyQ-mediated phenotypes (47, 48). At least for the ELF3-Col allele, however, our F1 data are not consistent with intragenic suppressors. Consistent with polyQ-mediated background effects, in at least one case, a modifier mutation has been shown to delay onset of Huntington's disease (11). Hypothetically, population genetic approaches could identify incompatible alleles that may contribute to variable disease onset in patients with polyQ expansions and to ELF3-dependent background effects in A. thaliana. However, the great diversity of TR alleles compared to SNP alleles and the small number of individuals carrying specific TR alleles render a population genetics approach infeasible. Extensive genetic mapping or other experimental approaches will be needed to identify the determinants of *ELF3-TR* dependent background effects. As TRs are rapidly evolving, we speculate that polyQ-mediated incompatibilities and the resulting fitness loss in hybrids and their offspring may contribute to disruption of gene flow between closely related populations. This speciation mechanism would be of particular importance for organisms with many polyQ-encoding TRs, thousands of offspring, and an inbreeding life style. Even in humans, however, about 1% of proteins contain polyQ tracts (13, 14, 45). Our results identify TR copy number variation, and in particular polyQ variation, as a phenotypically important class of genetic variation that warrants genome-wide assessment in model organisms, crops, and humans alike.

Chapter 3

THE CONSERVED *PFT1* TANDEM REPEAT IS CRUCIAL FOR PROPER FLOWERING IN *ARABIDOPSIS THALIANA*

A version of this chapter was published under the following reference:

Pauline Rival, Maximilian O. Press, Jacob Bale, Tanya Grancharova, Soledad F. Undurraga, and Christine Queitsch.The Conserved PFT1 Tandem Repeat is Crucial for Proper Flowering in *Arabidopsis thaliana*. *Genetics*, 198(2):747-754, August 2014.

Pauline Rival, Jacob Bale, Tanya Grancharova, and Soledad Undurraga generated transgenic lines and performed experiments.

Figures and tables prefixed with an 'S' can be found in Appendix B.

3.1 Summary

It is widely appreciated that short tandem repeat (STR) variation underlies substantial phenotypic variation in organisms. Some propose that the high mutation rates of STRs in functional genomic regions facilitate evolutionary adaptation. Despite their high mutation rate, some STRs show little to no variation in populations. One such STR occurs in the Arabidopsis thaliana gene PFT1 (MED25), where it encodes an interrupted polyglutamine tract. Though the PFT1 STR is large (270 bp), and thus expected to be extremely variable, it shows only minuscule variation across A. thaliana strains. We hypothesized that the PFT1 STR is under selective constraint, due to previously undescribed roles in PFT1 function. We investigated this hypothesis using plants expressing transgenic PFT1 constructs with either an endogenous STR or with synthetic STRs of varying length. Transgenic plants carrying the endogenous PFT1 STR gener-

ally performed best in complementing a pft1 null mutant across adult PFT1-dependent traits. In stark contrast, transgenic plants carrying a PFT1 transgene lacking the STR phenocopied a pft1 loss-of-function mutant for flowering time phenotypes, and were generally hypomorphic for other traits, establishing the functional importance of this domain. Transgenic plants carrying various synthetic constructs occupied the phenotypic space between wild-type and pft1-loss-of-function mutants. By varying PFT1 STR length, we discovered that PFT1 can act as either an activator or repressor of flowering in a photoperiod-dependent manner. We conclude that the PFT1 STR is constrained to its approximate wild-type length by its various functional requirements. Our study implies that there is strong selection on STRs not only to generate allelic diversity, but also to maintain certain lengths pursuant to optimal molecular function.

3.2 Introduction

Short tandem repeats (STRs, microsatellites) are ubiquitous and unstable genomic elements that have extremely high mutation rates (Subramanian et al. 2003; Legendre et al. 2007; Eckert and Hile 2009), leading to STR unit number variation within populations. STR variation in coding and regulatory regions can have significant phenotypic consequences (Gemayel et al. 2010). For example, several devastating human diseases, including Huntington□s disease and spinocerebellar ataxias, are caused by expanded STR alleles (Hannan 2010). However, STR variation can also confer beneficial phenotypic variation and may facilitate adaptation to new environments (Fondon et al. 2008; Gemayel et al. 2010). For example, in Saccharomyces cerevisiae natural polyQ variation in the FLO1 protein underlies variation in flocculation, which is important for stress resistance and biofilm formation in yeasts (Verstrepen et al. 2005). Natural STR variants of the Arabidopsis thaliana gene ELF3, which encode variable polyQ tracts, can phenocopy elf3 loss-of-function phenotypes in a common reference background (Undurraga et al. 2012). Moreover, the phenotypic effects of ELF3 STR variants differed dramatically between the divergent backgrounds Col and Ws, consistent with the exis-

tence of background-specific modifiers. Genetic incompatibilities involving variation in several other STRs have been described in plants, flies, and fish (Peixoto et al. 1998; Scarpino et al. 2013; Rosas et al. 2014). Taken together, these observations argue that STR variation underlies substantial phenotypic variation, and may also underlie some genetic incompatibilities. The A. thaliana gene PHYTOCHROME AND FLOWERING TIME 1 (PFT1, MEDIATOR 25, MED25) contains an STR of unknown function. In contrast to the comparatively short and pure ELF3 STR, the PFT1 STR encodes a long (90 amino acids in PFT1, vs. 7-29 for ELF3), periodically interrupted polyQ tract. The far greater length of the PFT1 STR leads to the prediction that its allelic variation should be greater than that of the highly variable ELF3 STR (Legendre et al. 2007, http://www.igs.cnrs-mrs.fr/TandemRepeat/Plant/index.php). However, in a set of diverse A. thaliana strains, PFT1 STR variation was negligible compared to that of the ELF3 STR (Supp. Table 1). Also, unlike ELF3, the PFT1 polyQ is conserved in plants as distant as rice, though its purity decreases with increasing evolutionary distance from A. thaliana. A glutamine-rich C-terminus is conserved even in metazoan MED25 (File S1). Recent studies of coding STRs suggested that there may be different classes of STR. Specifically, conserved tandem repeats appear in genes with substantially different functions from genes containing non-conserved tandem repeats (Schaper et al. 2014). Consequently, PFT1/MED25 polyQ conservation may functionally differentiate the PFT1 STR from the ELF3 STR. PFT1 encodes a subunit of Mediator, a conserved multi-subunit complex that acts as a molecular bridge between enhancer-bound transcriptional regulators and RNA polymerase II to initiate transcription (B□ckstr□m et al. 2007; Conaway and Conaway 2011). PFT1/MED25 is shared across multicellular organisms but absent in yeast. In A. thaliana, the PFT1 protein binds to at least 19 different transcription factors (Elfving et al. 2011; Ou et al. 2011; □evik et al. 2012; Chen et al. 2012) and has known roles in regulating a diverse set of processes such as organ size determination (Xu and Li 2011), ROS signaling in roots (Sundaravelpandian et al. 2013), biotic and abiotic stress (Elfving et al. 2011; Kidd et al. 2009; Chen et

al. 2012), phyB-mediated-light signaling, shade avoidance and flowering (Cerd□n and Chory 2003; Wollenberg et al. 2008; I□igo, Alvarez, et al. 2012; Klose et al. 2012). PFT1 was initially identified as a nuclear protein that negatively regulates the phyB pathway to promote flowering in response to specific light conditions (Cerd□n and Chory 2003; Wollenberg et al. 2008). Recently, I□igo and colleagues (2012) showed that PFT1 activates CONSTANS (CO) transcription and FLOWERING LOCUS T (FT) transcription in a CO-independent manner. Specifically, proteasome-dependent degradation of PFT1 is required to activate FT transcription and to promote flowering (I□igo, Giraldez, et al. 2012). The wide range of PFT1-dependent phenotypes is unsurprising given its function in transcription initiation, yet it remains poorly understood how PFT1 integrates these many signaling pathways. Given the conservation of the PFT1 polyQ tract and the known propensity of polyQ tracts for protein-protein and protein-DNA interactions (Escher et al. 2000; Schaefer et al. 2012), we hypothesized that this polyQ tract plays a role in the integration of multiple signaling pathways and is hence functionally constrained in length. We tested this hypothesis by generating transgenic lines expressing PFT1 with STRs of variable length and evaluating these lines for several PFT1-dependent developmental phenotypes. We show that the PFT1 STR is crucial for PFT1 function, and that PFT1-dependent phenotypes vary significantly with the length of the PFT1 STR. Specifically, the endogenous STR allele performed best for complementing the flowering and shade avoidance defects of the pft1-2 null mutant, though not for early seedling phenotypes. Our data indicate that most assayed PFT1-dependent phenotypes require a permissive PFT1 STR length. Taken together, our results suggest that the natural PFT1 STR length is constrained by the requirement of integrating multiple signaling pathways to determine diverse adult phenotypes.

3.3 Methods

3.3.1 Cloning

A 1000 bp region directly upstream of the PFT1 coding region was amplified and cloned into the pBGW gateway vector (Karimi et al. 2002) to create the entry vector pBGW-PFT1p. A full-length PFT1 cDNA clone, BX816858, was obtained from the French Plant Genomic Resources Center (INRA, CNRGV), and used as the starting material for all our constructs. The PFT1 gene was cloned into the pENTR4 gateway vector (Invitrogen) and the repeat region was modified by site-directed mutagenesis with QuikChange (Agilent Technologies), followed by restriction digestions and ligations. The modified PFT1 alleles were finally transferred to the pBGW-PFT1p vector via recombination using LR clonase (Invitrogen) to yield the final expression vectors. Seven constructs expressing various polyQ lengths (Table B2), plus an empty vector control, were used to transform homozygous pft1-2 mutants by the floral dip method (Clough and Bent 1998). Putative transgenics were selected for herbicide resistance with Basta (Liberty herbicide; Bayer Crop Science) and the presence of the transgene was confirmed by PCR analysis. Homozygous T3 and T4 plants with relative PFT1 expression levels between 0.5 and 4 times the expression of Col-0 were utilized for all experiments described. A minimum of two independent lines per construct was used for all experiments.

3.3.2 Expression analysis

All protocols were performed according to manufacture□s recommendations unless otherwise noted. Total RNA was extracted from 30mg of 10-days-old seedlings with the Promega SV Total RNA Isolation System (Promega). 2 ?g of total RNA were subjected to an exhaustive DNaseI treatment using the Ambion Turbo DNA-free Kit (Life Technologies). cDNA was synthesized from 100-300 ng of DNase-treated RNA samples with the Roche Transcriptor First Strand cDNA Synthesis Kit (Roche). Quanti-

tative Real-Time PCR was performed in a LightCyler®480 system (Roche) using the 480 DNA SYBR Green I Master kit. Three technical replicates were done for each sample. RT-PCR was performed under the following conditions: 5 min at 95 °C, followed by 35 cycles of 15 s at 95 °C, 20 s at 55 °C, and 20s at 72 °C. After amplification, a melting-curve analysis was performed. Expression of UBC21 (At5g25760) was measured as a reference in each sample, and used to calculate relative PFT1 expression. All expression values were normalized relative to WT expression, which was always set to 1.0. To measure splice forms, the protocol was the same but reactions were carried out in a standard thermal cycler and visualized on 2% agarose stained with ethidium bromide. For primers, see Table B4.

3.3.3 Plant Materials and Growth Conditions

Homozygous plants for the T-DNA insertional mutant SALK 129555, pft1-2, were isolated by PCR analysis from an F2 population obtained from the Arabidopsis Stock Center (ABRC) (Alonso et al. 2003). Plants were genotyped with the T-DNA specific primer LBb1 (`http://signal.salk.edu/tdna_FAQs.html`) and gene-specific primers (Table B4). Seeds were stratified at 40C for 3 days prior to shifting to the designated growth conditions, with the shift day considered day 0. For flowering time experiments, plants were seeded using a randomized design with 15-20 replicates per line in 4x9 pot trays. Trays were rotated 1800 and one position clockwise everyday in order to further reduce any possible position effect. Plants for LD were grown in 16 hours of light and 8 hours of darkness per 24 hour period. Bolting was called once the stem reached 1 cm in height. Full strength MS media containing MES, vitamins, 1% sucrose, and 0.24% phytagar was used for hypocotyl experiments. For germination experiments, half-strength MS media was used, supplemented with 1% sucrose, 0.5 g/L MES, and 2.4 g/L phytagel containing 200 mM NaCl or H_2O mock treatment with the pH adjusted to 5.7. All media was sterilized by autoclaving with 30 minutes of sterilization time. Seeds for tissue culture were surface sterilized with ethanol treatment prior to plat-

ing and left at 4oC for 3 days prior to shifting to the designated growth conditions. Plants for hypocotyl experiments were grown with 16 hours at 22 °C and 8 hours at 20oC in continuous darkness following an initial 2 hour exposure to light in order to induce germination. Germination experiments were scored on day 4 under LD at 20-22 °C. ImageJ software was utilized to make all hypocotyl and root length measurements. Raw phenotypic data are included as File S3.

3.3.4 Statistical analysis

All statistical analyses and plots were performed in R version 2.15.1 with ? = 0.05 (R Development Core Team 2012). Phenotypic data were analyzed using the analysis of variance (ANOVA), followed by Tukey□s HSD tests for the differences of groups within the ANOVA. Tukey□s HSD is a standard post-hoc test for multiple comparisons of the means of groups with homogeneous variance that corrects for the number of comparisons performed. Principal component analysis was performed using the prcomp() function after scaling each phenotypic variable to mean=0 and variance=1 across lines (phenotypes are not measured on the same quantitative scale; for example, SD flowering time ranges from 80 to 140 days, whereas LD rosette leaves ranges 5-15 leaves).

3.3.5 Sequence Analysis

Length of ELF3 and PFT1 STRs were determined by Sanger (dideoxy) sequencing. Raw sequencing data are available on the *Genetics* website (http://www.genetics.org/content/198/2/747.long). PFT1 and MED25 reference amino acid sequences were obtained from KEGG (Ogata et al. 1999) and aligned with Clustal Omega v1.0.3 with default options (Sievers et al. 2011).

3.4 Natural variation of PFT1 *STR*

We used Sanger sequencing to evaluate our expectation of high PFT1 STR variation across A. thaliana strains. However, we observed only three alleles of very similar size

(encoding 88, 89 and 90 amino acids, Table B1), in contrast to six different alleles of the much shorter ELF3 STR among these strains, some of which are three times the length of the reference allele (Undurraga et al. 2012). These data implied that the PFT1 and ELF3 STRs respond to different selective pressures. In coding STRs, high variation has been associated with positive selection (Laidlaw et al. 2007), though some basal level of neutral variation is expected due to the high mutation rate of STRs. We hypothesized that the PFT1 STR was constrained to this particular length by PFT1□s functional requirements. To test this hypothesis, we generated transgenic A. thaliana carrying PFT1 transgenes with various STR lengths in an isogenic pft1-2 mutant background. These transgenics included an empty vector control (VC), 0R, 0.34R, 0.5R, .75R, 1R (endogenous PFT1 STR allele), 1.27R, and 1.5R constructs. All STRs are given as their approximate proportion of WT STR length □ for instance, the 1R transgenic line contains the WT STR allele in the pft1-2 background (Table B2). We used expression analysis to select transgenic lines with similar PFT1 expression levels (Table B3).

3.5 The PFT1 STR length is essential for wild-type flowering and shade avoidance

We first evaluated the functionality of the different transgenic lines in flowering phenotypes. Removing the STR entirely substantially delayed flowering under long days (LD, phenotypes days to flower, rosette leaf number at flowering; Figure 1A). In LD, any STR allele other than 0R was able to rescue the pft1-2 late-flowering phenotype. Indeed, one allele (1.5R) showed earlier flowering than WT (Figure 1B, 1C), whereas other alleles provided a complete or nearly complete rescue of the pft1-2 mutant (Figure 1D). In short days (SD), we observed an unexpected reversal in rosette leaf phenotypes (compare SD and LD rosette leaves, Figures 1B, 1D). Rather than flowering late (adding more leaves) as in LD, the loss-of-function pft1-2 mutant appeared to flower early (fewer leaves at onset of flowering). Only the endogenous STR (1R) fully rescued this unexpected phenotype (Figure 1D). We observed the same mean trend for days to flowering

in SD, although differences were not statistically significant, even for pft1-2 (Figure 1D). This discrepancy may be due to insufficient power, or to a physiological decoupling of number of rosette leaves at flowering and days to flowering phenotypes in pft1-2 under SD conditions. Regardless, our results indicate that pft1-2□s late-flowering phenotype is specific to LD conditions. Our observation of this reversal in flowering time-related phenotypes appears to contradict previous data (Cerd□n and Chory 2003). However, a closer examination of this data reveals that the previously reported rosette leaf numbers in SD for the pft1-2 mutant show a similar trend. PFT1 STR length shows an approximately linear positive relationship with the SD rosette leaf phenotype, forming an allelic series of phenotypic severity. This allelic series strongly supports our observation of either slower growth rate (i.e. delayed addition of leaves) or early flowering of pft1-2 as measured by SD rosette leaves at flowering. PFT1 genetically interacts with the red/far-red light receptor phyB, which governs petiole length through the shade avoidance response (Cerd□n and Chory 2003; Wollenberg et al. 2008). We measured petiole length at bolting for plants grown under LD to evaluate the strength of their shade avoidance response, and thus whether the genetic interaction is affected by repeat length. Like the flowering time phenotypes, we found that the 1R allele most effectively rescued the long-petiole phenotype of the pft1-2 null among all STR alleles (Figure 2), though some alleles (e.g. 1.5R) show a rescue that is nearly as good. In summary, plants expressing the 1R transgene most closely resembled wild-type plants across a range of adult phenotypes. In contrast, the other STR alleles showed inconsistent performance across these phenotypes, rescuing only some phenotypes or at times out-performing wild-type.

3.6 PFT1 STR alleles fail to rescue early seedling phenotypes

We next assessed quantitative phenotypes in early seedling development, some of which had been previously connected to PFT1 function. Specifically, we measured hypocotyl and root length of dark-grown seedlings and examined germination in the presence of

salt (known to be defective in pft1 mutants) (Elfving et al. 2011). The pft1-2 mutant showed the previously reported effect on hypocotyl length as well as a novel defect in root length (Figure 3A). None of the transgenic lines, including the one containing the 1R allele, effectively rescued these pft1-2 phenotypes (Figure 3A). Similarly, 1R was not able to rescue the germination defect of pft1-2 on high-salt media. However, both the 1.5R and 0.5R alleles were able to rescue this phenotype (Figure 3B). In summary, no single STR allele, including the endogenous 1R, was consistently able to rescue the early seedling phenotypes of the pft1-2 mutant. One explanation for the failure of the endogenous STR (PFT1-1R) to rescue early seedling phenotypes is that the PFT1 transgene represents only the larger of two splice forms. The smaller PFT1 splice form, which we did not test, may play a more important role in early seedling development. To explore this hypothesis, we measured mRNA levels of the two splice forms in pooled 7-day seedlings grown under the tested conditions and various adult tissues at flowering in Col-0 plants. However, we found that both splice forms were expressed in all samples, and in all samples the larger splice form was the predominant form (data not shown). The possibility remains that downstream regulation or tissue-specific expression may lead to a requirement for the smaller splice form in early seedlings.

3.7 Summarizing PFT1 STR function across all tested phenotypes

Given the complex phenotypic responses to PFT1 STR substitutions, results were equivocal as to which STR allele demonstrated the most □wild-type-like□ phenotype across traits, as measured by its sufficiency in rescuing pft1-2 null phenotypes. To summarize the various phenotypes, we calculated the mean of each quantitative phenotype for each allele, and used principal component analysis (PCA) to visualize the joint distribution of phenotypes observed. All STR alleles were distributed between the pft1-2 null and wild-type (WT) in PC1, which was strongly associated with adult traits and represented a majority of phenotypic variation among lines (Figure 4). PC1 showed that 1R was the most generally efficacious allele for adult phenotypes. However, 1R showed

incomplete rescue in early seedling phenotypes such as hypocotyl length, which drove PC2. All STR alleles showed substantial rescue in adult phenotypes, and even the 0R allele showed a partial rescue in some phenotypes; however, rescue of early seedling phenotypes was generally poor for all alleles. The first principal component also captured our observation that the pft1-2 flowering defect reversed sign in SD vs. LD: according to Figure 4, SD and LD quantitative phenotypes are both strongly represented on principal component 1, but they show opposite directionality. We take this observation as support of this hitherto-unknown complexity in PFT1 function.

3.8 Discussion

STR-containing proteins pose an intriguing puzzle □they are prone to in-frame mutations, which in many instances lead to dramatic phenotypic changes (Gemayel et al. 2010). Although STR-dependent variation has been linked to adaptation in a few cases, the presence of mutationally labile STRs in functionally important core components of cell biology seems counterintuitive. PFT1, also known as MED25, is a core component of the transcriptional machinery across eukaryotes and contains an STR that is predicted to be highly variable in length. Contrary to this prediction, we found PFT1 STR variation to be minimal, consistent with substantial functional constraint. The existing residual variation (2% of reference STR length, as opposed to >100% for the ELF3 STR in the same A. thaliana strains) suggests that the PFT1 STR is mutationally labile like other STRs. In fact, several of the synthetic PFT1 alleles examined in this study arose spontaneously during cloning. Strong functional constraint, however, may select against such deviations in STR length in planta. Here, we establish the essentiality of the full-length PFT1 STR and its encoded polyQ tract for proper PFT1 function in A. thaliana. We found that diverse developmental phenotypes were altered by the substitution of alternative STR lengths for the endogenous length. Leveraging the support of the PFT1 STR allelic series, we report new aspects of PFT1 function in flowering time and root development.

3.8.1 The PFT1 STR is required for PFT1 function in adult traits

The PFT1 0R lines did not effectively complement pft1-2 for adult phenotypes, suggesting a crucial role of the PFT1 STR in regulating the onset of flowering and shade avoidance. Generally, PFT1-1R was most effective in producing wild-type-like adult phenotypes. The precise length of the STR, however, seemed less important for the onset of flowering in LD. With exception of PFT1-0R, all other STR alleles were also able to rescue the loss-of-function mutant to some extent, suggesting that as long as some repeat sequence is present, the PFT1 gene product can fulfill this function. Under other conditions, and for other adult phenotypes, requirements for PFT1 STR length appeared more stringent. Specifically, under SD, the rosette leaf number phenotype of the pft1-2 mutant can only be rescued by PFT1-1R, while STR alleles perform worse with increasing distance from this length □optimum□.

3.8.2 pft1-2 mutants are late-flowering in LD but not SD

pft1-2 plants had fewer rosette leaves at flowering in SD, but more rosette leaves in LD, consistent with previous, largely undiscussed observations (Cerd□n and Chory 2003). Under LD conditions, pft1 null mutants flowered late, as described in several previous studies (Cerd□n and Chory 2003; Wollenberg et al. 2008), but we observe no such phenotype under SD conditions, contradicting at least one prior study (Cerd□n and Chory 2003). These data suggest that while PFT1 functions as a flowering activator under LD, its role is more complex under SD. One recent study showed that PFT1 function in LD is dependent upon its ability to bind E3 ubiquitin ligases (I□igo, Giraldez, et al. 2012). Inhibition of proteasome activity also prevents PFT1 from promoting FT transcription and thus inducing flowering, suggesting that degradation of PFT1 or associated proteins is a critical feature of PFT1□s transcriptional activation of flowering in LD. If this degradation is somehow down-regulated in SD, PFT1 could switch from a flowering activator to a repressor, through decreased Mediator complex turnover at

promoters. Recent studies raised the possibility that different PFT1-dependent signaling cascades have different requirements for PFT1 turnover (Ou et al. 2011; Kidd et al. 2009), which may contribute to the condition-specific PFT1 flowering phenotype we observe. Conservatively, we conclude that the regulatory process that mediates the phenotypic reversal between LD and SD depends on the endogenous PFT1 STR allele, suggesting that the polyQ is crucial to PFT1□s activity as both activator and potentially as a repressor of flowering.

3.8.3 Incomplete complementation of germination and hypocotyl length by the PFT1 constructs

Whereas pft1-2 adult phenotypes were rescued by the PFT1-1R allele, most of our transgenic lines could not fully rescue pft1-2 early seedling phenotypes of 1) germination under salt, 2) hypocotyl length, and 3) root length. The PFT1 gene is predicted to have two different splice forms, the larger of which was used to generate our constructs (both splice forms contain the STR). Several studies have shown that, under stress conditions, different splice forms of the same gene can play distinct roles (Yan et al. 2012; Leviatan et al. 2013; Staiger and Brown 2013). We note that the conditions under which PFT1-1R fails to complement are also potentially stressful conditions (artificial media, sucrose, high salt, dark). The shorter splice form of PFT1 may be required in signaling pathways triggered under stress conditions. We presume that the failure to complement results from a deficiency related to this missing splice form. However, hypocotyl length was the only trait in which all examined STR alleles resembled the pft1-2 mutant. The significant functional differentiation among the STR alleles for root length and germination suggests that the large splice form does retain at least some function in early seedling traits.

3.8.4 Implications for STR and PFT1 biology

Implications for STR and PFT1 biology: Coding and regulatory STRs have been previously studied and discussed as a means of facilitating evolutionary innovation (Verstrepen et al. 2005). However, this means of innovation is based upon the same sequence characteristics that promote protein-protein and protein-DNA binding (Escher et al. 2000; Schaefer et al. 2012), such that STR variability must be balanced against functional constraints. This balance has recently been described for a set of 18 coding dinucleotide STRs in humans, which are maintained by natural selection even though any mutation is likely to cause frame-shift mutations (Haasl and Payseur 2014). These results, coupled with our observations, lend credence to these authors previous argument that not all STRs act as agents of adaptive change (Haasl and Payseur 2013). Considering again the possibility that more conserved coding tandem repeats have distinct functions from non-conserved tandem repeats (Schaper et al. 2014), we suggest that PFT1 and ELF3 can serve as models for these two selective regimes, and that the structural roles of their respective polyQs underlie the differences in natural variation between the two. In some cases, such as ELF3, high variability is not always inconsistent with function, even while holding genetic background constant (Undurraga et al. 2012). In PFT1, we have identified a STR whose low variability reflects strong functional constraints. We speculate that these constraints are associated with a structural role for the PFT1 polyQ in the Mediator complex, either in protein-protein interactions with other subunits or in protein-DNA interactions with target promoters. Given that a glutamine-rich C-terminus appears to be a conserved feature of MED25 even in metazoans (File S1), we expect that our results are generalizable to Mediator function wherever this protein is present. Future work will be necessary in understanding possible mechanisms by which the MED25 polyQ might facilitate Mediator complex function and contribute to ontogeny throughout life. Moreover, attempts must be made to understand the biological and structural characteristics unique to polyQ-

containing proteins that tolerate (or encourage) polyQ variation, as opposed to those polyQ-containing proteins (like PFT1) that are under strong functional constraints.

Chapter 4

SHORT TANDEM REPEATS AND QUANTITATIVE GENETICS

Portions of this chapter were published under the following references:

- Maximilian O. Press, Keisha D. Carlson, and Christine Queitsch.

  The overdue promise of short tandem repeat variation for heritability. *Trends in Genetics*, 30(11):504-512, August 2014.

- Keisha D. Carlson, Peter H. Sudmant, Maximilian O. Press, Evan E. Eichler, Jay Shendure, and Christine Queitsch. MIPSTR: a method for multiplex genotyping of germline and somatic STR variation across many individuals. *Genome Research*, 25(5):750-761, May 2015.

Figures and tables prefixed with an 'S' can be found in Appendix C.

Chapter 5

THE VARIABLE ELF3 POLYGLUTAMINE HUBS AN EPISTATIC NETWORK

Figures and tables prefixed with an 'S' can be found in Appendix D.

Chapter 6

ELF3 POLYGLUTAMINE VARIATION REVEALS A PIF4-INDEPENDENT ROLE IN THERMORESPONSIVE FLOWERING

Figures and tables prefixed with an 'S' can be found in Appendix E.

Chapter 7

GENOME-SCALE CO-EVOLUTIONARY INFERENCE IDENTIFIES FUNCTIONS AND CLIENTS OF BACTERIAL HSP90

A version of this chapter was published under the following reference:

Maximilian O. Press, Hui Li, Nicole Creanza, Guenter Kramer, Christine Queitsch, Victor Sourjik, and Elhanan Borenstein. Genome-scale co-evolutionary inference identifies functions and clients of bacterial Hsp90. *PLoS Genetics*, 9(7):e1003631, 2013.

Figures and tables prefixed with an 'S' can be found in Appendix F.

Chapter 8

EVOLUTIONARY ASSEMBLY PATTERNS OF PROKARYOTIC GENOMES

A version of this chapter is under review for publication, and is available at http://biorxiv.org/conte

Figures and tables prefixed with an 'S' can be found in Appendix G.

Chapter 9
THE THESIS UNFORMATTED

This chapter describes the uwthesis class (`uwthesis.cls`, version dated 2011/06/27) in detail and shows how it was used to format the thesis. A working knowledge of Lamport's LaTeX manual[**?**] is assumed.

9.1 The Control File

The source to this sample thesis is contained in a single file only because ease of distribution was a concern. You should not do this. Your task will be much easier if you break your thesis into several files: a file for the preliminary pages, a file for each chapter, one for the glossary, and one for each appendix. Then use a control file to tie them all together. This way you can edit and format parts of your thesis much more efficiently.

Figure 9.1 shows a control file that might have produced this thesis. It sets the document style, with options and parameters, and formats the various parts of the thesis—but contains no text of its own.

The first section, from the `\documentclass` to the `\begin\{document\}`, defines the document class and options. This thesis has specified two-sided formatting, which is now allowed by the Graduate School. Two sided printing is now actually LaTeX's default. If you want one sided printing you must specify `oneside`. This sample also specified a font size of 11 points. Possible font size options are: `10pt`, `11pt`, and `12pt`. Default is 12 points, which is the preference of the Graduate School. If you choose a smaller size be sure to check with the Graduate School for acceptability. The smaller fonts can produce very small sub and superscripts.

Include most additional formatting packages with `\usepackage`, as describe by Lamport[**?**].

Figure 9.1: A thesis control file (`thesis.tex`). This file is the input to LaTeX that will produce a thesis. It contains no text, only commands which direct the formatting of the thesis. This is also an example of a 'facing page' caption. It is guaranteed to appear on a lefthand page, facing the figure contents on the right. See the text.

```
% LaTeX thesis control file

\documentclass[11pt,twoside]{uwthesis}

\begin{document}

% preliminary pages
%
\prelimpages
\include{prelim}

% text pages
%
\textpages
\include{chap1}
\include{chap2}
\include{chap3}
\include{chap4}

% bibliography
%
\bibliographystyle{plain}
\bibliography{all}

% appendices
%
\appendix
\include{appxa}
\include{appxb}

\include{vita}
\end{document}
```

The one exception to this rule is the `natbib` package. Include it with the `natbib` document option.

Use the `\includeonly` command to format only a part of your thesis. See Lamport[**?**, sec. 4.4] for usage and limitations.

9.2 The Text Pages

A chapter is a major division of the thesis. Each chapter begins on a new page and has a Table of Contents entry.

9.2.1 Chapters, Sections, Subsections, and Appendices

Within the chapter title use a `\\` control sequence to separate lines in the printed title (recall Figure **??**.). The `\\` does not affect the Table of Contents entry.

Format appendices just like chapters. The control sequence `\appendix` instructs LaTeX to begin using the term 'Appendix' rather than 'Chapter'.

Sections and subsections of a chapter are specified by `\section` and `\subsection`, respectively. In this thesis chapter and section titles are written to the table of contents. Consult Lamport[**?**, pg. 176] to see which subdivisions of the thesis can be written to the table of contents. The `\\` control sequence is not permitted in section and subsection titles.

9.2.2 Footnotes

Footnotes format as described in the LaTeX book. You can also ask for end-of-chapter or end-of-thesis notes. The thesis class will automatically set these up if you ask for the document class option `chapternotes` or `endnotes`.

If selected, chapternotes will print automatically. If you choose endnotes however you must explicitly indicate when to print the notes with the command `\printendnotes`. See the style guide for suitable endnote placement.

9.2.3 Figures and Tables

Standard LaTeX figures and tables, see Lamport[**?**, sec. C.9], normally provide the most convenient means to position the figure. Full page floats and facing captions are exceptions to this rule.

If you want a figure or table to occupy a full page enclose the contents in a `fullpage` environment. See figures 9.2.

Facing page captions are described in the Style Manual[**?**]. They have different meanings depending on whether you are using the one-side or two-side thesis style.

If you are using the two-side style, facing captions are full page captions for full page figures or tables and must face the illustration to which they refer. You must explicitly format both pages. The caption part must appear on an even page (left side) and the figure or table must come on the following odd page (right side). Enclose the float contents for the caption in a `leftfullpage` environment, and enclose the float contents for the figure or table in a `fullpage` environment. Figure 9.1, for example, required a full page so its caption (on a facing caption page) would have been formatted as shown in figure 9.2a. The first page (left side) contains the caption. The second page (right side) could be left blank. A picture or graph might be pasted onto this space.

If instead you are using the one-side style, facing caption pages are still captions for full page figures or tables that appear on the left-hand page (facing the illustration on the right-hand page). However, the page number and binding offset are reversed from their normal positions. Format these captions by enclosing the float contents in a `leftfullpage` environment. Because you are printing on only one side of each sheet, you must manually turn over this caption sheet. You then have the choice of inserting a preprinted illustration or formatting one to print with the thesis. In either case no page number should appear on the illustration page, nor should the page number increment. Enclose your figure's text in an `xtrafullpage` environment, which will cause the page numbers to come out right. You can, of course, leave out the illustration and insert a

```
\begin{figure}[p]% the left side caption
  \begin{leftfullpage}
    \caption{ . . . }
  \end{leftfullpage}
\end{figure}
\begin{figure}[p]% the right side space
  \begin{fullpage}
     . . .
     ( note.. no caption here )
  \end{fullpage}
\end{figure}
```

Figure 9.2: (

a)This text would create a double page figure in the two-side style.

preprinted copy later. Figure 9.2b shows how to format a facing caption page in the one-side style. Note that, in this case, the illustration was also printed.

In the two-side style the `xtrafullpage` environment acts just like the `fullpage` environment. It does not produce a numberless page.

9.2.4 Horizontal Figures and Tables

Figures and tables may be formatted horizontally (a.k.a. landscape) as long as their captions appear horizontal also. LaTeX will format landscape material for you if a couple of conditions are met. You have to have a printer and printer driver that allow rotations and you have to have a couple of add-on LaTeX packages.

Include the `rotating` package

```
\usepackage[figuresright]{rotating}
```

and read the documentation that comes with the package.

Figure 9.4 is an example of how a landscape table might be formatted.

```
\begin{figure}[p]
   \begin{leftfullpage}
      \caption{ . . . }
   \end{leftfullpage}
\end{figure}
\begin{figure}[p]% the right side space
  \begin{xtrafullpage}
     . . .
     ( note.. no caption here )
  \end{xtrafullpage}
\end{figure}
```

Figure 9.3: (

b)[Generating a facing caption page]This text would create a facing caption page with the accompaning figure in the one-side style.

```
\begin{sidewaystable}
    ...
    \caption{ . . . }
\end{sidewaystable}
```

Figure 9.4: This text would create a landscape table with caption.

9.2.5 Figure and Table Captions

Most captions are formatted with the `\caption` macro as described by Lamport[**?**, sec. C.9]. The uwthesis class extends this macro to allow continued figures and tables, and to provide multiple figures and tables with the same number, e.g., 3.1a, 3.1b, etc.

To format the caption for the first part of a figure or table that cannot fit onto a single page use the standard form:

```
\caption[toc]{text}
```

To format the caption for the subsequent parts of the figure or table use this caption:

```
\caption(-){(continued)}
```

It will keep the same number and the text of the caption will be *(continued)*.

To format the caption for the first part of a multi-part figure or table use the format:

```
\caption(a)[toc]{text}
```

The figure or table will be lettered (with 'a') as well as numbered. To format the caption for the subsequent parts of the multi-part figure or table use the format:

```
\caption(x){text}
```

where *x* is `b`, `c`, ... The parts will be lettered (with 'b', 'c', ...).

9.3 The Preliminary Pages

These are easy to format only because they are relatively invariant among theses. Therefore the difficulties have already been encountered and overcome by LaTeX and the thesis document classes.

Start with the definitions that describe your thesis. This sample thesis was printed with the parameters:

```
\Title{The Suitability of the \LaTeX\ Text Formatter\\
   for Thesis Preparation by Technical and\\
   Non-technical Degree Candidates}
\Author{Jim Fox}
\Program{UW Information Technology}
\Year{2012}

\Chair{Name of Chairperson}{title}{Chair's department}
\Signature{First committee member}
```

```
\Signature{Next committee member}
\Signature{etc}
```

Use two or more `\Chair` lines if you have co-chairs.

9.3.1 Copyright page

Print the copyright page with `\copyrightpage`.

9.3.2 Title page

Print the title page with `\titlepage`. The title page of this thesis was printed with[1]

```
\titlepage
```

You may change default text on the title page with these macros. You will have to redefine `$\degree$text`, for instance, if you're writing a Master's thesis instead of a dissertation.[2]

`\Degree{`*degree name*`}` defaults to "Doctor of Philosophy"

`\School{`*school name*`}` defaults to "University of Washington"

`\Degreetext{`*degree text*`}` defaults to "A dissertation submitted ..."

`\textofCommittee{`*committee label*`}` defaults to "Reading Committee:"

`\textofChair{`*chair label*`}` defaults to "Chair of the Supervisory Committee:"

These definitions must appear before the `\titlepage` command.

---

[1] Actually, it wasn't. I added a footnote—something you would not do.

[2] If you use these they can be included with the other information before copyrightpage".

9.3.3 Abstract

Print the abstract with `\abstract`. It has one argument, which is the text of the abstract. All the names have already been defined. The abstract of this thesis was printed with

```
\abstract{This sample . . . ‘real’ dissertation.}
```

9.3.4 Tables of contents

Use the standard LaTeX commands to format these items.

9.3.5 Acknowledgments

Use the `\acknowledgments` macro to format the acknowledgments page. It has one argument, which is the text of the acknowledgment. The acknowledgments of this thesis was printed with

```
\acknowledgments{The author wishes . . . {\it il miglior fabbro}.\par}}
```

BIBLIOGRAPHY

[1] Jennifer R Gatchel and Huda Y Zoghbi. Diseases of unstable repeat expansion: mechanisms and common principles. *Nature reviews. Genetics*, 6(10):743–55, October 2005.

[2] Rita Gemayel, Marcelo D Vinces, Matthieu Legendre, and Kevin J Verstrepen. Variable tandem repeats accelerate evolution of coding and regulatory sequences. *Annual review of genetics*, 44:445–77, January 2010.

[3] Harry T. Orr. Polyglutamine neurodegeneration: Expanded glutamines enhance native functions. *Current Opinion in Genetics and Development*, 22(3):251–255, 2012.

[4] William C Wimsatt. The Units of Selection and the Structure of the Multi-Level Genome. *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, 1980:122–183, January 1980.

Appendix A

SUPPORTING CHAPTER 2

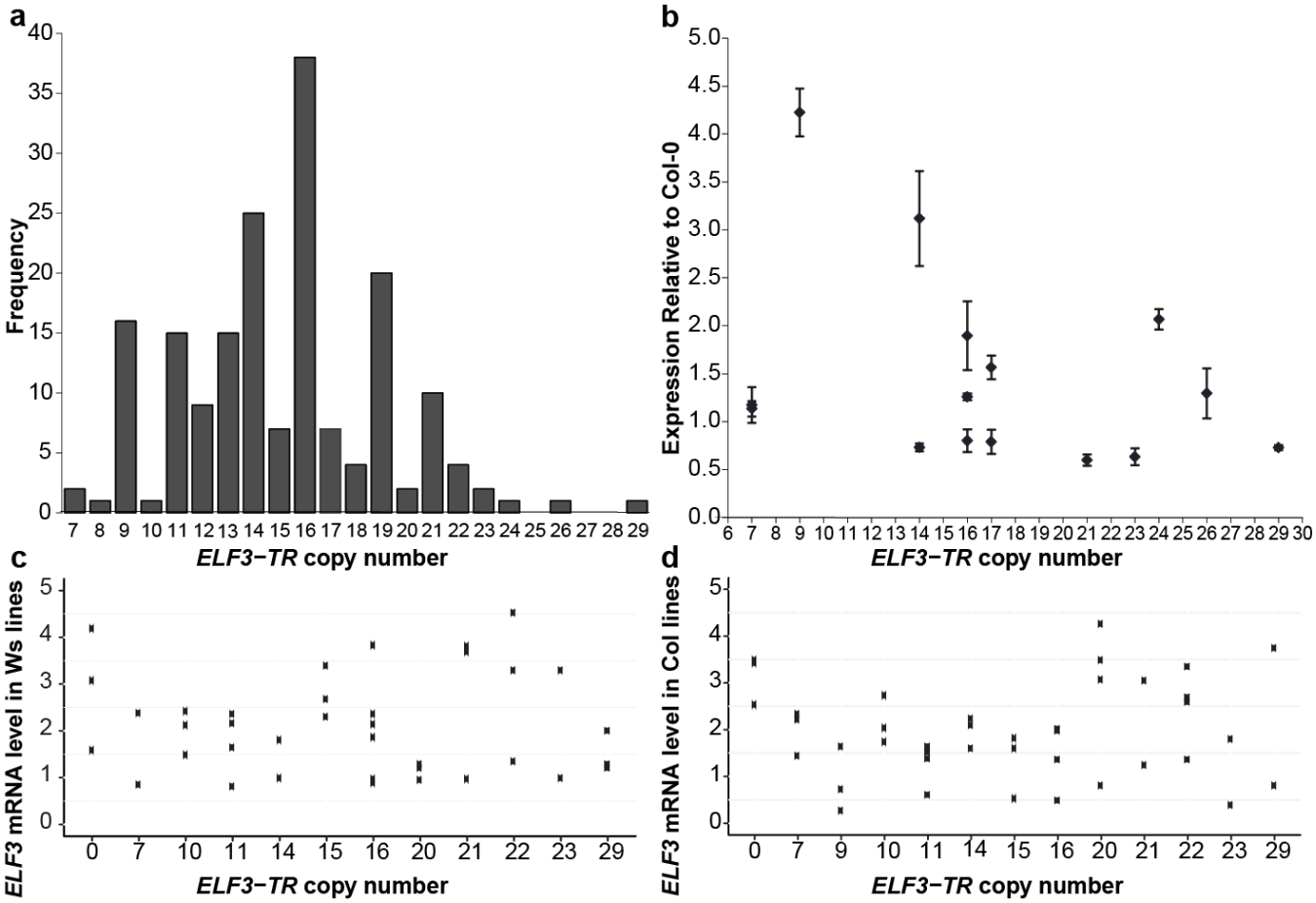
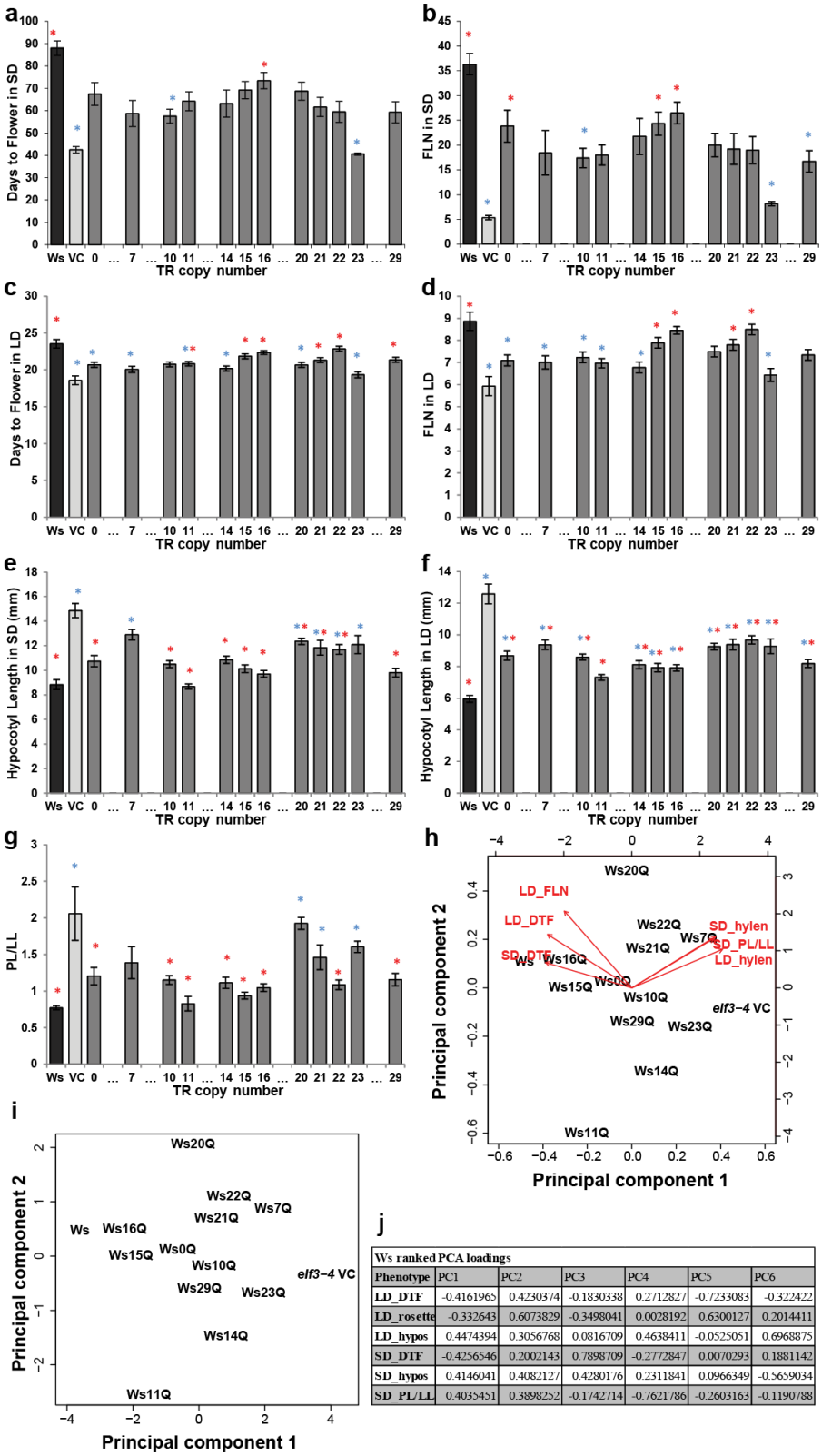

a
Frequency
ELF3−TR copy number
b
Expression Relative to Col-0
ELF3−TR copy number
c
ELF3 mRNA level in Ws lines
ELF3−TR copy number
d
ELF3 mRNA level in Col lines
ELF3−TR copy number

Figure A.1: The ELF3-TR variation is not correlated with ELF3 expression. (A) Histogram of ELF3-TR copy number across 181 accessions. TR copy number was determined by Sanger sequencing. (B) ELF3 expression levels in selected natural accessions were measured by quantitative RT-PCR. Expression values are given relative to the Col-0 wild-type reference. Three biological replicates with three technical replicates each were used to obtain expression values. Bars indicate ± SEM. (C and D) ELF3-TR transgenic lines are expression-matched in both genetic backgrounds. (C) elf3-4, Ws; (D) elf3-200, Col. ELF3 mRNA levels were measured by quantitative PCR (for primers see Table S6) in pooled 10-d-old seedlings that were grown under LD and collected at ZT 20 for each independently generated ELF3-TR transgenic line. ELF3 expression levels are shown relative to either Ws (C) or Col-0 (D) wild-types. Because ELF3 expression levels are known to substantially affect ELF3-dependent phenotypes [**?**], ELF3 expression is an important variable to consider in our assessment of polyQ tract-length effects. We made efforts to consider only lines within a certain range of ELF3 expression and to test multiple independent lines per ELF3-TR allele (Tables A2□A4), but because of the technical constraints of transgenic plant construction, we cannot entirely exclude the possibility that ELF3 expression partially explains our observations. Although the effects of both ELF3 expression level and ELF3-TR copy number were highly significant, they appear to be largely independent. For example, the ELF3-23Q and ELF3-16Q alleles, which were among the most distinct ELF3-TR alleles in both backgrounds, had very similar ranges of ELF3 expression. In Ws, the alleles ELF3-7Q, ELF3-23Q, and ELF3-10Q phenocopied an elf3 loss-of-function mutant for some phenotypes. Their ELF3 expression levels, however, were very similar to the ELF3-16Q allele, which complemented many ELF3 functions in elf3-4. As observed with individual ELF3-TR alleles, the phenotypic effects of ELF3 expression levels appear to be largely independent of ELF3-TR copy number, which consistently explained a larger portion of phenotypic variation.



j

Ws ranked PCA loadings						
Phenotype	PC1	PC2	PC3	PC4	PC5	PC6
LD_DTF	-0.4161965	0.4230374	-0.1830338	0.2712827	-0.7233083	-0.322422
LD_rosette	-0.332643	0.6073829	-0.3498041	0.0028192	0.6300127	0.2014411
LD_hypos	0.4474394	0.3056768	0.0816709	0.4638411	-0.0525051	0.6968875
SD_DTF	-0.4256546	0.2002143	0.7898709	-0.2772847	0.0070293	0.1881142
SD_hypos	0.4146041	0.4082127	0.4280176	0.2311841	0.0966349	-0.5659034
SD_PL/LL	0.4035451	0.3898252	-0.1742714	-0.7621786	-0.2603163	-0.1190788

Figure A.2: ELF3-TR variation has nonlinear phenotypic effects in the elf3-4 background (Ws accession). (A) Days to flower (DTF) under SD (n = 6 plants per line). (B) Final number of rosette leaves (FLN) under SD (n = 6 plants per line). (C) DTF under LD (n = 15 plants per line). (D) FLN under LD (n = 15 plants per line). (E) Hypocotyl length under SD (n = 20-30 seedlings per line). (F) Hypocotyl length under LD (n = 20-30 seedlings per line). (G) PL/LL ratio under SD (n = 6 plants per line). Data are from the same plants as in B. ELF3-TR alleles are indicated with the number of Qs encoded, Ws is wild-type, VC is the elf3-4 vector control. Blue and red asterisks indicate alleles that are significantly different from the wild-type and from the VC, respectively, by Tukey-HSD test ($\alpha = 0.05$). We used this analysis rather than the one presented in Figure 1B to preserve clarity. Bars indicate $\pm$ SEM. These experiments were repeated at least once with similar results. (H and I) PCA of phenotypic data for all ELF3-TR alleles in the elf3-4 background (Ws accession). (H) Biplot of PC1 and PC2, graphically showing the contribution of phenotypes to PCs as red arrows. Note that for the biplot representation, PC1 and PC2 are transformed to the same scale (bottom and left axes), whereas phenotype contributions (in red) are allowed to differ in scale (top and right axes). Phenotypes are hypocotyl length in short and long days (SD hylen and LD hylen), DTF in short and long days (SD DTF and LD DTF), and FLN in long days (LD rosette). Wild-type plants are characterized by late flowering (large SD and LD DTF, many rosette leaves) and short hypocotyls (small SD and LD hylen), relative to elf3 loss-of-function mutants. (I) PC1 and PC2. (J) PCA loadings for Ws background. hylen, hypocotyl length (mm). PCA loadings describe the composition/loading of each principal component . For PC1, flowering-time phenotypes and circadian clock phenotypes have opposite loading signs.

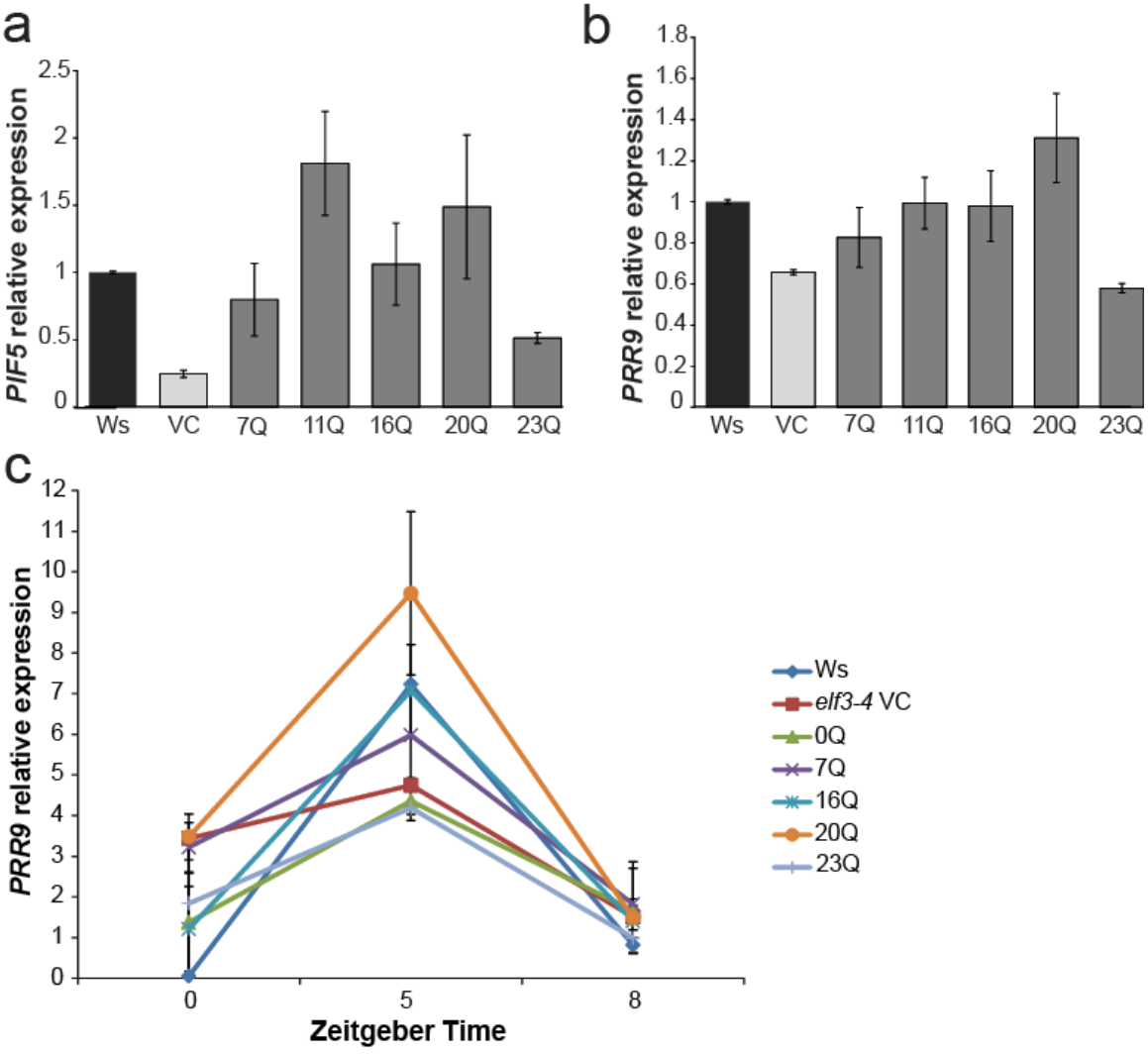

a
PIF5 relative expression
2.5
2
1.5
1
0.5
0
Ws
VC
7Q
11Q
16Q
20Q
23Q
b
PRR9 relative expression
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
Ws
VC
7Q
11Q
16Q
20Q
23Q
c
PRR9 relative expression
12
11
10
9
8
7
6
5
4
3
2
1
0
0
5
8
Zeitgeber Time
Ws
elf3-4 VC
0Q
7Q
16Q
20Q
23Q

Figure A.3: Expression levels of the ELF3-regulated genes PIF5 (A) and PRR9 (B and C). Plants were grown under LD and RNA was collected at times showing the largest expression difference between wild-type and elf3-4 mutant ZT8 for PIF5 [**?**] (A) and ZT5 for PRR9 [**?**, **?**] (B and C). RNA levels were normalized relative to Ws wild-type. (C) Temporal variation in PRR9 expression across ELF3-TR transgenic lines. PRR9 expression levels were measured in 10-d-old plants grown under LD. RNA was collected at times demonstrating the diurnal oscillation of PRR9 expression in wild-type, as determined previously. RNA levels were normalized relative to wild-type (Ws) at ZT8. Gene expression was measured in triplicate for each biological replicate, with multiple independent transgenic lines as biological replicates for each ELF3 allele. Error bars indicate SE of expression across biological replicates. Our expression patterns of PRR9 for wild-type and the elf3-4 mutant are similar to previous observations [**?**, **?**]. ELF3-TR alleles are indicated with the number of Qs encoded, Ws is wild-type, VC is the elf3-4 vector control. Error bars are SEs of means. Data are from multiple independently generated expression-matched (Figure A1C) T3 and T4 lines for each TR copy number allele (Table A2).

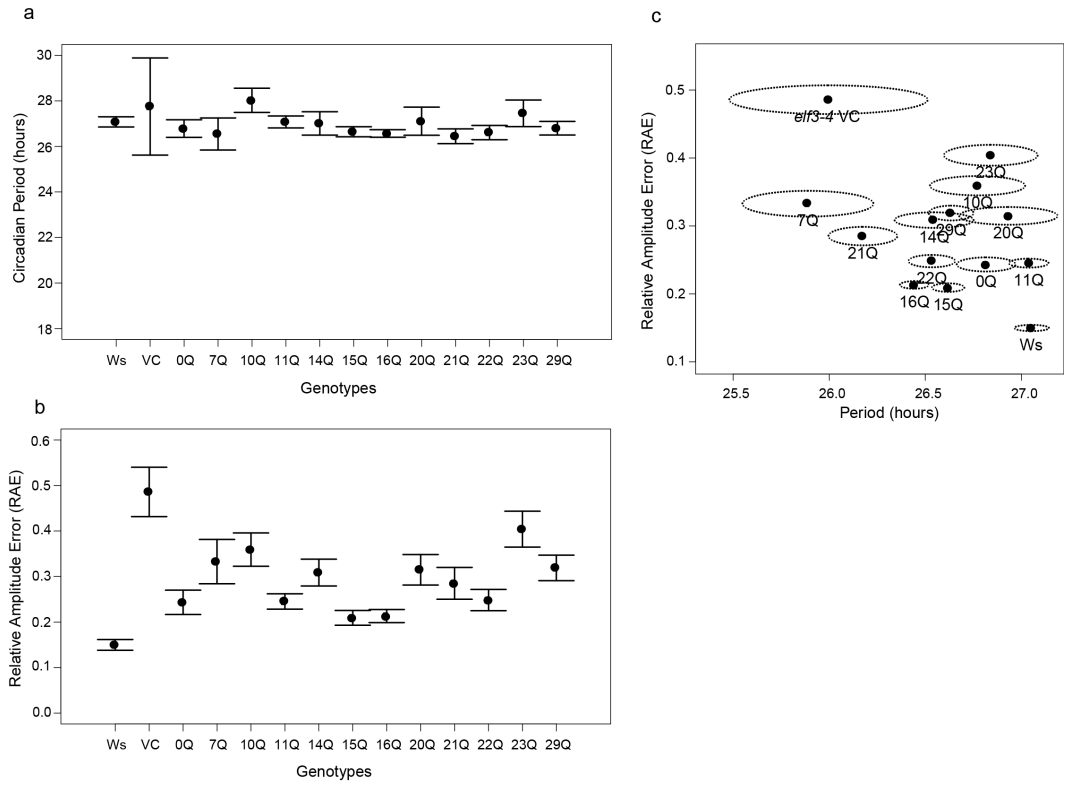
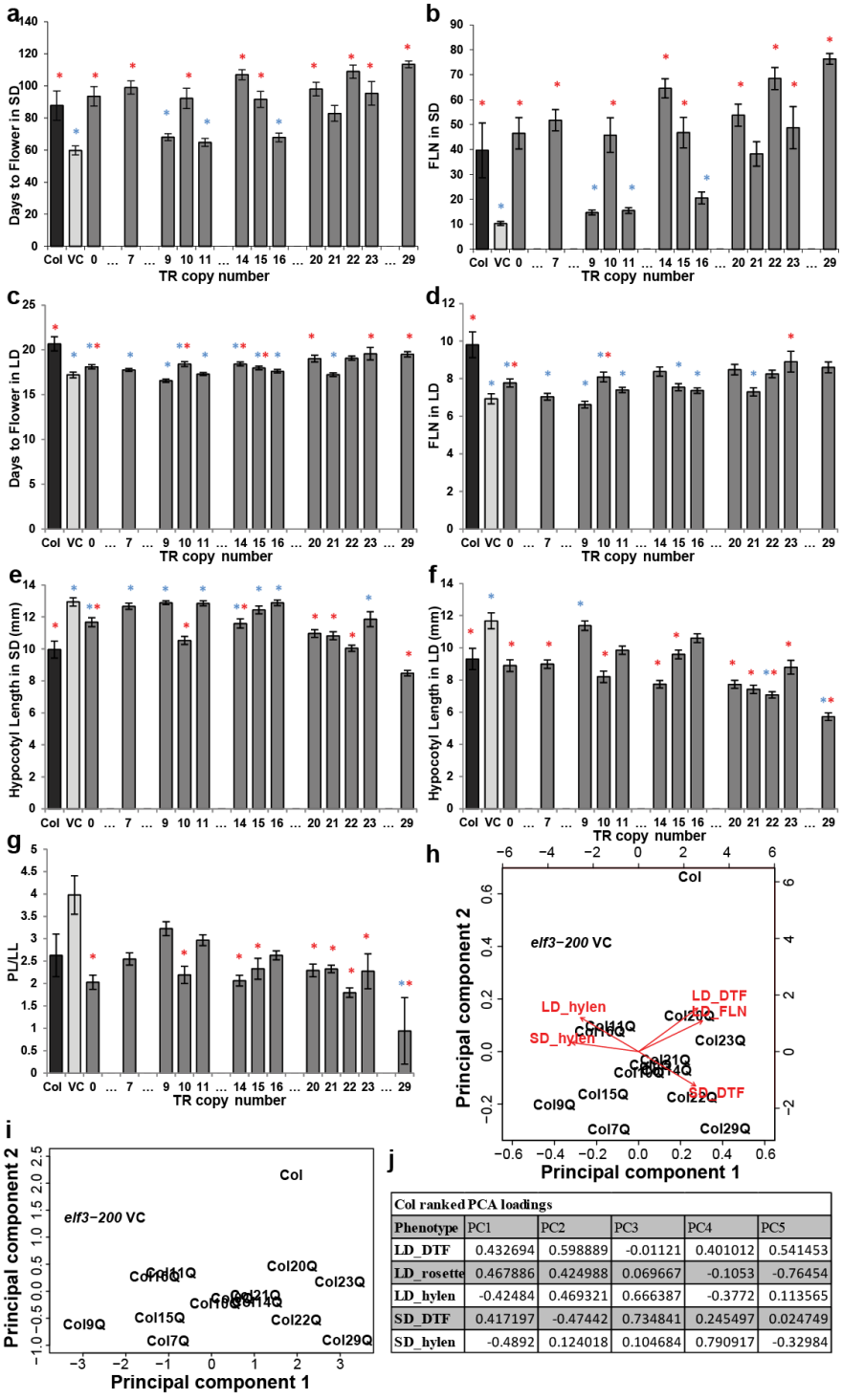
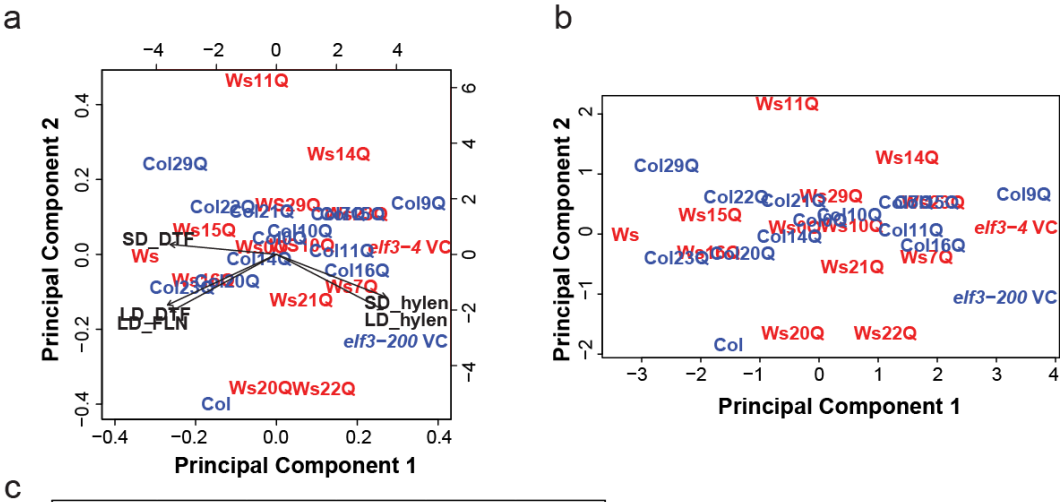


Figure A.4: Circadian parameters estimated for different TR alleles in elf3-4 CCR2::Luc reporter lines. (A) Measured circadian period of CCR::LUC expression oscillation for each ELF3-TR allele. Bars correspond to 99% confidence intervals for this proportion. (B) Measured RAE of CCR::LUC expression oscillation for each ELF3-TR allele. Bars correspond to 99% confidence intervals for this proportion. Plants with RAE <0.4 are considered to have a robust circadian clock. (C) Estimated RAE and circadian period for each ELF3-TR allele. Points are means, dotted ellipses represent SEMs, and genotype labels indicate ELF3-TR copy number. Bioluminescence of the CCR2::LUC reporter present in ELF3-TR transgenic lines was used to measure circadian parameters (period and RAE). Seedlings were entrained in 12-h light:12-h dark cycles for 5 d and released to LL on the sixth day. Note that plants with high RAE have by definition unreliable estimates of circadian period. Number of seedlings for each genotype: Ws, 274; 0Q, 249; 7Q, 122; 10Q, 222; 11Q, 339; 14Q, 214; 15Q, 284; 16Q, 534; 20Q, 161; 21Q, 243; 22Q, 271; 23Q, 196; 29Q, 257; elf3-4 vector control, 102.



Col ranked PCA loadings					
Phenotype	PC1	PC2	PC3	PC4	PC5
LD_DTF	0.432694	0.598889	-0.01121	0.401012	0.541453
LD_rosette	0.467886	0.424988	0.069667	-0.1053	-0.76454
LD_hylen	-0.42484	0.469321	0.666387	-0.3772	0.113565
SD_DTF	0.417197	-0.47442	0.734841	0.245497	0.024749
SD_hylen	-0.4892	0.124018	0.104684	0.790917	-0.32984

Figure A.5: ELF3-TR variation has nonlinear phenotypic effects in the elf3-200 background (Col-0 accession). (A) DTF under SD (n = 9 plants/line). (B) FLN under SD (n = 9 plants per line). (C) DTF under LD (n = 15 plants per line). (D) FLN under LD (n = 15 plants per line). (E) Hypocotyl length under SD (n = 20□30 seedlings per line). (F) Hypocotyl length under LD (n = 20□30 seedlings per line). (G) PL/LL ratio under SD (n = 9 plants per line). Data are from the same plants as in B. ELF3-TR alleles are indicated with the number of Qs encoded, Col is wild-type, VC is the elf3-200 vector control (VC). Blue and red asterisks indicate alleles that are significantly different from the wild-type and from the vector control, respectively, by Tukey-HSD test (α = 0.05). Bars indicate $\pm$ SEM. These experiments were repeated at least once with similar results. (H and I) PCA of phenotypic data for all ELF3-TR alleles in the elf3-200 (Col accession) background. (H) Biplot of PC1 and PC2, graphically showing the contribution of phenotypes to PCs as red arrows. Note that for the biplot representation, PC1 and PC2 are transformed to the same scale (bottom and left axes), whereas phenotype contributions (in red) are allowed to differ in scale (top and right axes). Phenotypes are hypocotyl length in short and long days (SD hylen and SD hylen), DTF in short and long days (SD DTF and SD DTF), and FLN in long days (SD FLN). Wild-type type plants are characterized by late flowering (large SD and SD DTF, many rosette leaves) and short hypocotyls (small SD and SD hylen), relative to elf3 loss-of-function mutants. (I) PC1 and PC2. Note that PC1□s orientation is inverted relative to PCAs including Ws-background plants (A and B: i.e., elf3-200 is to the negative end of the axis, and Col is at the positive end); this does not affect interpretation. In contrast to PCAs including Ws data, PC2 of Col data alone represents the differential response of LD and SD phenotypes to ELF3-polyQ copy number variation. (J) PCA loadings for Col background. hylen = hypocotyl length (mm).



c

Ws and Col backgrounds ranked PCA					
Phenotype	PC1	PC2	PC3	PC4	PC5
LD_DTF	-0.45341	-0.48687	0.21643	-0.06427	0.711617
LD_rosette	-0.44167	-0.5515	0.097755	-0.08292	-0.69595
LD_hylen	0.442085	-0.52945	-0.16168	0.705025	0.032283
SD_DTF	-0.44214	0.09598	-0.87758	0.145134	0.063971
SD_hylen	0.456536	-0.41143	-0.38381	-0.6862	0.064153

Figure A.6: The phenotypic effects of ELF3-TR copy number variation are strongly background-dependent. PCA of phenotypic data from all ELF3-TR alleles in both elf3-4 (Ws accession) and elf3-200 (Col accession) backgrounds. (A) Biplot of PC1 and PC2, graphically showing the contribution of phenotypes to PCs as black arrows. Note that for the biplot representation, PC1 and PC2 are transformed to the same scale (bottom and left axes), whereas phenotype contributions (in red) are allowed to differ in scale (top and right axes). Phenotypes are hypocotyl length in short and long days (SD hylen and LD hylen), DTF in short and long days (SD DTF and LD DTF), and FLN in long days (LD FLN). Wild-type plants are characterized by late flowering (large SD and LD DTF, many rosette leaves) and short hypocotyls (small SD and LD hylen), relative to elf3 loss-of-function mutants. Text in red represents a given allele in the Ws background (transgenics in elf3-4), and text in blue represents alleles in the Col background (transgenics in elf3-200). (B) PC1 and PC2. (C) PCA loadings for both back- grounds. hylen = hypocotyl length (mm).

Appendix B

SUPPORTING CHAPTER 3

Appendix C

SUPPORTING CHAPTER 4

Appendix D

SUPPORTING CHAPTER 5

Appendix E

SUPPORTING CHAPTER 6

Appendix F

SUPPORTING CHAPTER 7

Appendix G

WHERE TO FIND THE FILES

The uwthesis class file, `uwthesis.cls`, contains the parameter settings.